

An operational characterization of the notion of probability by algorithmic randomness and its applications*

Kohtaro Tadaki

Department of Computer Science, College of Engineering, Chubu University
1200 Matsumoto-cho, Kasugai-shi, Aichi 487-8501, Japan

E-mail: tadaki@cs.chubu.ac.jp

<http://www2.odn.ne.jp/tadaki/>

Abstract. The notion of probability plays an important role in almost all areas of science and technology. In modern mathematics, however, probability theory means nothing other than measure theory, and the operational characterization of the notion of probability is not established yet. In this paper, based on the toolkit of algorithmic randomness we present an operational characterization of the notion of probability, called an *ensemble*. Algorithmic randomness, also known as algorithmic information theory, is a field of mathematics which enables us to consider the randomness of an individual infinite sequence. We use the notion of Martin-Löf randomness with respect to Bernoulli measure to present the operational characterization. As the first step of the research of this line, in this paper we consider the case of finite probability space, i.e., the case where the sample space of the underlying probability space is finite, for simplicity. We give a natural operational characterization of the notion of conditional probability in terms of ensemble, and give equivalent characterizations of the notion of independence between two events based on it. Furthermore, we give equivalent characterizations of the notion of independence of an arbitrary number of events/random variables in terms of ensembles. In particular, we show that the independence between events/random variables is equivalent to the independence in the sense of van Lambalgen’s Theorem, in the case where the underlying finite probability space is computable. In the paper we make applications of our framework to information theory and cryptography in order to demonstrate the wide applicability of our framework to the general areas of science and technology.

Key words: probability, algorithmic randomness, operational characterization, Martin-Löf randomness, Bernoulli measure, conditional probability, independence, van Lambalgen’s Theorem, information theory, cryptography

*This work combines two preliminary papers, entitled: “An operational characterization of the notion of probability by algorithmic randomness,” which appeared in the Proceedings of the 37th Symposium on Information Theory and its Applications (SITA2014), 5.4.1, pp. 389–394, December 9–12, 2014, Unazuki, Toyama, Japan, and: “An operational characterization of the notion of probability by algorithmic randomness and its application to cryptography,” which appeared in the Proceedings of the 32nd Symposium on Cryptography and Information Security (SCIS2015), 2D4-3, January 20–23, 2015, Kokura, Japan.

1 Introduction

The notion of probability plays an important role in almost all areas of science and technology. In modern mathematics, however, probability theory means nothing other than *measure theory*, and the operational characterization of the notion of probability is not established yet. In this paper, based on the toolkit of *algorithmic randomness* we present an operational characterization of the notion of probability. Algorithmic randomness is a field of mathematics which enables us to consider the randomness of an individual infinite sequence. We use the notion of *Martin-Löf randomness with respect to Bernoulli measure* to present the operational characterization.

To clarify our motivation and standpoint, and the meaning of the operational characterization, let us consider a familiar example of a probabilistic phenomenon. We here consider the repeated throwings of a fair die. In this probabilistic phenomenon, as throwings progressed, a specific infinite sequence such as

$$3, 5, 6, 3, 4, 2, 2, 3, 6, 1, 5, 3, 5, 4, 1, \dots$$

is being generated, where each number is the outcome of the corresponding throwing of the die. Then the following naive question may arise naturally.

Question: What property should this infinite sequence satisfy as a probabilistic phenomenon?

In this paper we try to answer this question. We characterize the notion of probability as an infinite sequence of outcomes in a probabilistic phenomenon of a *specific mathematical property*. We call such an infinite sequence of outcomes the *operational characterization of the notion of probability*. As the specific mathematical property, in this paper we adopt the notion of *Martin-Löf randomness with respect to Bernoulli measure*, a notion in algorithmic randomness.

We put forward this proposal as a thesis (see Thesis 1 in Section 5). We check the validity of the thesis based on our intuitive understanding of the notion of probability. Furthermore, we characterize *equivalently* the basic notions in probability theory in terms of the operational characterization. Namely, we equivalently characterize the notion of the *independence* of random variables/events in terms of the operational characterization, and represent the notion of *conditional probability* in terms of the operational characterization in a natural way. The existence of these equivalent characterizations confirms further the validity of the thesis.

1.1 Historical background

In the past century, there was a comprehensive attempt to provide an operational characterization of the notion of probability. Namely, von Mises developed a mathematical theory of repetitive events which was aimed at reformulating the theory of probability and statistics based on an operational characterization of the notion of probability [30, 31]. In a series of his comprehensive works which began in 1919, von Mises developed this theory and, in particular, introduced the notion of *collective* as a mathematical idealization of a long sequence of outcomes of experiments or observations repeated under a set of invariable conditions, such as the repeated tossings of a coin or of a pair of dice.

The collective plays a role as an operational characterization of the notion of probability, and is an infinite sequence of sample points in the sample space of a probability space. As the randomness property of the collective, von Mises assumes that all “reasonable” infinite subsequences of a

collective satisfy the law of large numbers with the identical limit value, where the subsequences are selected using “acceptable selection rules.” Wald [32, 33] later showed that for any countable collection of selection rules, there are sequences which are collectives in the sense of von Mises. However, at the time it was unclear exactly what types of selection rules should be acceptable. There seemed to von Mises to be no canonical choice.

Later, with the development of computability theory and the introduction of generally accepted precise mathematical definitions of the notions of algorithm and computable function, Church [7] suggested that a selection rule be considered acceptable if and only if it is computable. In 1939, however, Ville [29] revealed the defect of the notion of collective. Namely, he showed that for any countable collection of selection rules, there is a sequence that is random in the sense of von Mises but has properties that make it clearly nonrandom. In the first place, the collective has an *intrinsic defect* that it cannot exclude the possibility that an event with probability zero may occur. (For the development of the theory of collectives from the point of view of the definition of randomness, see Downey and Hirschfeldt [9].)

In 1966, Martin-Löf [14] introduced the definition of random sequences, which is called *Martin-Löf randomness* nowadays, and plays a central role in the recent development of algorithmic randomness. At the same time, he introduced the notion of *Martin-Löf randomness with respect to Bernoulli measure* [14]. He then pointed out that this notion overcomes the defect of the collective in the sense of von Mises, and this can be regarded precisely as the collective which von Mises wanted to define. However, he did not develop probability theory based on Martin-Löf random sequence with respect to Bernoulli measure.

Algorithmic randomness is a field of mathematics which studies the definitions of random sequences and their property (see [16, 9] for the recent developments of the field). However, the recent research on algorithmic randomness would seem only interested in the notions of randomness themselves and their interrelation, and not seem to have made an attempt to develop probability theory based on Martin-Löf randomness with respect to Bernoulli measure in an operational manner so far.

1.2 Contribution of the paper

The subject of this paper is to make such an attempt. Namely, in this paper we present an operational characterization of the notion of probability based on Martin-Löf randomness with respect to Bernoulli measure. We call it an *ensemble*, instead of collective for distinction. The name “ensemble” comes from physics, in particular, from quantum mechanics and statistical mechanics. We propose to identify it with an infinite sequence of outcomes resulting from the infinitely repeated trials in a probabilistic phenomenon. We show that the ensemble has enough properties to regard it as an operational characterization of the notion of probability from the point of view of our intuitive understanding of the notion of probability.

Actually, we give a natural operational characterization of the notion of conditional probability in terms of ensemble, and give equivalent characterizations of the notion of independence between two events based on it. Furthermore, we give equivalent characterizations of the notion of independence of an arbitrary number of events/random variables in terms of ensembles. In particular, we show that the independence of events/random variables is equivalent to the independence in the sense of van Lambalgen’s Theorem [28], in the case where the underlying probability space is *computable*.

As the first step of the research of this line, in this paper we consider only the case of *finite probability space*, i.e., the case where the sample space of the underlying probability space is finite, for simplicity. The investigation of the case of general probability spaces is reported in the sequels to the paper.

We emphasize that the Bernoulli measure which we consider in this paper is not required to be computable at all (except for the results related to van Lambalgen's Theorem), while the measures considered in algorithmic randomness so far are usually computable. Thus, the central results in this paper hold for any finite probability space.

Finally, we make applications of our framework to information theory and cryptography as examples of the fields for the applications, in order to demonstrate the wide applicability of our framework to the general areas of science and technology.

Modern probability theory originated from the *axiomatic approach* to probability theory, introduced by Kolmogorov [13] in 1933, where the probability theory is precisely *measure theory*. One of the important roles of modern probability theory is, of course, in its applications to the general areas of science and technology. As we have already pointed out, however, an operational characterization of the notion of probability is still missing in modern probability theory. Thus, when we apply the results of modern probability theory, we have no choice but to make such applications *thoroughly based on our intuition without formal means*.

The aim of this paper is to try to fill in this gap between modern probability theory and its applications. We present the operational characterization of the notion of probability as a *rigorous interface* between theory and practice, without appealing to our intuition for filling in the gap. Anyway, in this work we *keep* modern probability theory *in its original form* without any modifications, and propose the operational characterization of the notion of probability as an *additional mathematical structure* to it, which provides modern probability theory with more comprehensive and rigorous opportunities for applications.

1.3 Organization of the paper

The paper is organized as follows. We begin in Section 2 with some preliminaries to measure theory, computability theory, and algorithmic randomness. In Section 3, we introduce the notion of finite probability space on which the operational characterization of the notion of probability is presented. Based on the notion of finite probability space we then introduce the notion of Martin-Löf randomness with respect to Bernoulli measure in Section 4.

In Section 5 we introduce the notion of ensemble, and put forward a thesis which states to identify the ensemble as an operational characterization of the notion of probability. We then check the validity of the thesis. In Section 6 we start to construct our framework by characterizing operationally the notions of conditional probability and the independence between two events, in terms of ensembles. We then characterize operationally the notion of the independence of an arbitrary number of events/random variables in terms of ensembles in Section 7. In Section 8 we show that the independence notions, introduced in the preceding sections, are further equivalent to the notion of the independence in the sense of van Lambalgen's Theorem, in the case where the underlying finite probability space is computable, by generalizing van Lambalgen's Theorem over our framework. Thus we show that the three independence notions, considered in this paper, are all equivalent in this case.

In Section 9 we make applications of our framework to information theory and cryptography.

We mention an application to quantum mechanics there. We conclude this paper with a mention of the next step of the research in Section 10.

2 Preliminaries

2.1 Basic notation and definitions

We start with some notation about numbers and strings which will be used in this paper. $\#S$ is the cardinality of S for any set S . $\mathbb{N} = \{0, 1, 2, 3, \dots\}$ is the set of *natural numbers*, and \mathbb{N}^+ is the set of *positive integers*. \mathbb{Q} is the set of *rational numbers*, and \mathbb{R} is the set of *reals*.

An *alphabet* is a nonempty finite set. Let Ω be an arbitrary alphabet throughout the rest of this section. A *finite string over Ω* is a finite sequence of elements from the alphabet Ω . We use Ω^* to denote the set of all finite strings over Ω , which contains the *empty string* denoted by λ . We use Ω^+ to denote the set $\Omega^* \setminus \{\lambda\}$. For any $\sigma \in \Omega^*$, $|\sigma|$ is the *length* of σ . Therefore $|\lambda| = 0$. For any $\sigma \in \Omega^+$ and $k \in \mathbb{N}^+$ with $k \leq |\sigma|$, we use $\sigma(k)$ to denote the k th element in σ . Therefore, we have $\sigma = \sigma(1)\sigma(2)\dots\sigma(|\sigma|)$ for every $\sigma \in \Omega^+$. For any $n \in \mathbb{N}$, we use Ω^n and $\Omega^{\leq n}$ to denote the sets $\{x \mid x \in \Omega^* \text{ \& } |x| = n\}$ and $\{x \mid x \in \Omega^* \text{ \& } |x| \leq n\}$, respectively. A subset S of Ω^* is called *prefix-free* if no string in S is a prefix of another string in S .

An *infinite sequence over Ω* is an infinite sequence of elements from the alphabet Ω , where the sequence is infinite to the right but finite to the left. We use Ω^∞ to denote the set of all infinite sequences over Ω . Let $\alpha \in \Omega^\infty$. For any $n \in \mathbb{N}$, we denote by $\alpha|_n \in \Omega^*$ the first n elements in the infinite sequence α and by $\alpha(n)$ the n th element in α . Thus, for example, $\alpha|_4 = \alpha(1)\alpha(2)\alpha(3)\alpha(4)$, and $\alpha|_0 = \lambda$. For any $S \subset \Omega^*$, the set $\{\alpha \in \Omega^\infty \mid \exists n \in \mathbb{N} \alpha|_n \in S\}$ is denoted by $[S]^\prec$. Note that (i) $[S]^\prec \subset [T]^\prec$ for every $S \subset T \subset \Omega^*$, and (ii) for every set $S \subset \Omega^*$ there exists a prefix-free set $P \subset \Omega^*$ such that $[S]^\prec = [P]^\prec$. For any $\sigma \in \Omega^*$, we denote by $[\sigma]^\prec$ the set $[\{\sigma\}]^\prec$, i.e., the set of all infinite sequences over Ω extending σ . Therefore $[\lambda]^\prec = \Omega^\infty$.

2.2 Measure theory

We briefly review measure theory according to Nies [16, Section 1.9]. See also Billingsley [4] for measure theory in general.

A real-valued function μ defined on the class of all subsets of Ω^∞ is called an *outer measure on Ω^∞* if the following conditions hold.

- (i) $\mu(\emptyset) = 0$;
- (ii) $\mu(\mathcal{C}) \leq \mu(\mathcal{D})$ for every subsets \mathcal{C} and \mathcal{D} of Ω^∞ with $\mathcal{C} \subset \mathcal{D}$;
- (iii) $\mu(\bigcup_i \mathcal{C}_i) \leq \sum_i \mu(\mathcal{C}_i)$ for every sequence $\{\mathcal{C}_i\}_{i \in \mathbb{N}}$ of subsets of Ω^∞ .

A *probability measure representation over Ω* is a function $r: \Omega^* \rightarrow [0, 1]$ such that

- (i) $r(\lambda) = 1$ and
- (ii) for every $\sigma \in \Omega^*$ it holds that

$$r(\sigma) = \sum_{a \in \Omega} r(\sigma a). \quad (1)$$

A probability measure representation r over Ω induces an outer measure μ_r on Ω^∞ in the following manner: A subset \mathcal{R} of Ω^∞ is *open* if $\mathcal{R} = [S]^\prec$ for some $S \subset \Omega^*$. Let r be an arbitrary probability measure representation over Ω . For each open subset \mathcal{A} of Ω^∞ , we define $\mu_r(\mathcal{A})$ by

$$\mu_r(\mathcal{A}) := \sum_{\sigma \in E} r(\sigma),$$

where E is a prefix-free subset of Ω^* with $[E]^\prec = \mathcal{A}$. Due to the equality (1), the sum is independent of the choice of the prefix-free set E and therefore the value $\mu_r(\mathcal{A})$ is well-defined. Then, for any subset \mathcal{C} of Ω^∞ , we define $\mu_r(\mathcal{C})$ by

$$\mu_r(\mathcal{C}) := \inf\{\mu_r(\mathcal{A}) \mid \mathcal{C} \subset \mathcal{A} \text{ \& } \mathcal{A} \text{ is an open subset of } \Omega^\infty\}.$$

We can then show that μ_r is an *outer measure* on Ω^∞ such that $\mu_r(\Omega^\infty) = 1$.

A class \mathcal{F} of subsets of Ω^∞ is called a σ -*field* on Ω^∞ if \mathcal{F} includes Ω^∞ , is closed under complements, and is closed under the formation of countable unions. The *Borel class* \mathcal{B}_Ω is the σ -field *generated by* all open sets on Ω^∞ . Namely, the Borel class \mathcal{B}_Ω is defined as the intersection of all the σ -fields on Ω^∞ containing all open sets on Ω^∞ . A real-valued function μ defined on the Borel class \mathcal{B}_Ω is called a *probability measure on Ω^∞* if the following conditions hold.

- (i) $\mu(\emptyset) = 0$ and $\mu(\Omega^\infty) = 1$;
- (ii) $\mu(\bigcup_i \mathcal{D}_i) = \sum_i \mu(\mathcal{D}_i)$ for every sequence $\{\mathcal{D}_i\}_{i \in \mathbb{N}}$ of sets in \mathcal{B}_Ω such that $\mathcal{D}_i \cap \mathcal{D}_j = \emptyset$ for all $i \neq j$.

Then, for every probability measure representation r over Ω , we can show that the restriction of the outer measure μ_r on Ω^∞ to the Borel class \mathcal{B}_Ω is a probability measure on Ω^∞ . We denote the restriction of μ_r to \mathcal{B}_Ω by μ_r just the same.

Then it is easy to see that

$$\mu_r([\sigma]^\prec) = r(\sigma) \tag{2}$$

for every probability measure representation r over Ω and every $\sigma \in \Omega^*$.

2.3 Computability

A partial function $f: \mathbb{N} \rightarrow \Omega^*$ or $f: \mathbb{N} \rightarrow \mathbb{Q}$ is called *partial computable* if there exists a deterministic Turing machine \mathcal{M} such that, for each $n \in \mathbb{N}$, when executing \mathcal{M} with the input n ,

- (i) if $n \in \text{dom } f$ then the computation of \mathcal{M} eventually terminates and then \mathcal{M} outputs $f(n)$;
- (ii) if $n \notin \text{dom } f$ then the computation of \mathcal{M} does not terminate,

where $\text{dom } f$ denotes the domain of the definition of f . A function $f: \mathbb{N} \rightarrow \Omega^*$ or $f: \mathbb{N} \rightarrow \mathbb{Q}$ is called *computable* if f is partial computable with $\text{dom } f = \mathbb{N}$. A computable function is also called a *total recursive function*. We say that $\alpha \in \Omega^\infty$ is *computable* if the mapping $\mathbb{N} \ni n \mapsto \alpha|_n$ is a computable function.

A real a is called *computable* if there exists a computable function $g: \mathbb{N} \rightarrow \mathbb{Q}$ such that $|a - g(k)| < 2^{-k}$ for all $k \in \mathbb{N}$. A real a is called *left-computable* if there exists a computable, increasing sequence of rationals which converges to a , i.e., if there exists a computable function

$h: \mathbb{N} \rightarrow \mathbb{Q}$ such that $h(n) \leq h(n+1)$ for every $n \in \mathbb{N}$ and $\lim_{n \rightarrow \infty} h(n) = a$. On the other hand, a real a is called *right-computable* if $-a$ is left-computable. It is then easy to see that, for every $a \in \mathbb{R}$, a is computable if and only if a is both left-computable and right-computable.

A subset \mathcal{C} of $\mathbb{N}^+ \times \Omega^*$ is called *recursively enumerable* (r.e., for short) if there exists a deterministic Turing machine \mathcal{M} such that, for each $x \in \mathbb{N}^+ \times \Omega^*$, when executing \mathcal{M} with the input x ,

- (i) if $x \in \mathcal{C}$ then the computation of \mathcal{M} eventually terminates;
- (ii) if $x \notin \mathcal{C}$ then the computation of \mathcal{M} does not terminate.

2.4 Algorithmic randomness

In the following we concisely review some definitions and results of algorithmic randomness [5, 6, 16, 9].

We use \mathcal{L} to denote Lebesgue measure on $\{0, 1\}^\infty$. Namely, $\mathcal{L} = \mu_r$ where r is a probability measure representation over $\{0, 1\}^\infty$ defined by the condition that $r(\sigma) = 2^{-|\sigma|}$ for every $\sigma \in \{0, 1\}^*$. The idea in algorithmic randomness is to think of an infinite binary sequence as *random* if it is in no *effective null set*. An effective null set is a subset \mathcal{S} of $\{0, 1\}^\infty$ such that $\mathcal{L}(\mathcal{S}) = 0$ and \mathcal{S} has some type of *effective* property. To specify an algorithmic randomness notion, one has to specify a type of effective null set, which is usually done by introducing a test concept. Failing the test is the same as being in the null set. In this manner, various randomness notions, such as 2-randomness, weak 2-randomness, Demuth randomness, Martin-Löf randomness, Schnorr randomness, and Kurtz randomness, have been introduced so far, and a hierarchy of algorithmic randomness notions has been developed (see [16, 9] for the detail of the hierarchy).

Among all randomness notions, *Martin-Löf randomness* is a central one. This is because in many respects, Martin-Löf randomness is well-behaved, in that the many properties of Martin-Löf random infinite sequences do match our intuition of what random infinite sequence should look like. Moreover, the concept of Martin-Löf randomness is robust in the sense that it admits various equivalent definitions which are all natural and intuitively meaningful, as we will partly see in Theorem 2. Martin-Löf randomness is defined as follows, based on the notion of *Martin-Löf test*.

Definition 1 (Martin-Löf randomness, Martin-Löf [14]). *A subset \mathcal{C} of $\mathbb{N}^+ \times \{0, 1\}^*$ is called a Martin-Löf test if \mathcal{C} is an r.e. set, and for every $n \in \mathbb{N}^+$ it holds that*

$$\mathcal{L}([C_n]^\prec) < 2^{-n}$$

where \mathcal{C}_n denotes the set $\{\sigma \mid (n, \sigma) \in \mathcal{C}\}$.

For any $\alpha \in \{0, 1\}^\infty$, we say that α is Martin-Löf random if for every Martin-Löf test \mathcal{C} there exists $n \in \mathbb{N}^+$ such that $\alpha \notin [C_n]^\prec$. \square

Let \mathcal{C} be a Martin-Löf test. Then, for each $k \in \mathbb{N}^+$, we see that

$$\mathcal{L}\left(\bigcap_{n=1}^{\infty} [C_n]^\prec\right) \leq \mathcal{L}([C_k]^\prec) < 2^{-k}.$$

On letting $k \rightarrow \infty$, we have

$$\mathcal{L}\left(\bigcap_{n=1}^{\infty} [C_n]^\prec\right) = 0.$$

Thus, the set $\bigcap_{n=1}^{\infty} [\mathcal{C}_n]^{\prec}$ forms an effective null set in the notion of Martin-Löf randomness. Definition 1 says that an infinite binary sequence α is Martin-Löf random if α is not in the effective null set $\bigcap_{n=1}^{\infty} [\mathcal{C}_n]^{\prec}$ for any Martin-Löf test \mathcal{C} .

The robustness of Martin-Löf randomness is mainly due to the fact that it admits characterizations based on the notion of program-size complexity, as shown in Theorem 2. The *program-size complexity* (or *Kolmogorov complexity*) $K(\sigma)$ of a finite binary string σ is defined as the length of the shortest binary input for a universal decoding algorithm U , called an *optimal prefix-free machine*, to output σ (see Chaitin [5] for the detail). By the definition, $K(\sigma)$ can be thought of as the randomness contained in the individual finite binary string σ .

Theorem 2 (Schnorr [18], Chaitin [5]). *For every $\alpha \in \{0, 1\}^{\infty}$, the following conditions are equivalent:*

(i) α is Martin-Löf random.

(ii) There exists $c \in \mathbb{N}$ such that, for all $n \in \mathbb{N}^+$, $n - c \leq K(\alpha \upharpoonright_n)$. □

The condition (ii) means that the infinite binary sequence α is *incompressible*.

3 Finite probability spaces

In this paper we give an operational characterization of the notion of probability for a *finite probability space*. A finite probability space is formally defined as follows.

Definition 3. *Let Ω be an alphabet. A finite probability space on Ω is a function $P: \Omega \rightarrow \mathbb{R}$ such that*

(i) $P(a) \geq 0$ for every $a \in \Omega$, and

(ii) $\sum_{a \in \Omega} P(a) = 1$.

The set of all finite probability spaces on Ω is denoted by $\mathbb{P}(\Omega)$.

Let $P \in \mathbb{P}(\Omega)$. The set Ω is called the sample space of P , and elements of Ω are called sample points or elementary events of P . For each $A \subset \Omega$, we define $P(A)$ by

$$P(A) := \sum_{a \in A} P(a).$$

A subset of Ω is called an event on P , and $P(A)$ is called the probability of A for every event A on P . □

Let Ω be an arbitrary alphabet through out the rest of this paper. It plays a role of the set of all possible outcomes of experiments or observations. An operational characterization of the notion of probability which we give for a finite probability space on Ω is an infinite sequence over Ω .

*It is convenient to introduce the notion of *computable finite probability space* as follows.*

Definition 4. *Let $P \in \mathbb{P}(\Omega)$. We say that P is computable if $P(a)$ is a computable real for every $a \in \Omega$. □*

We may try to weaken the notion of the computability for a finite probability space as follows: Let $P \in \mathbb{P}(\Omega)$. We say that P is *left-computable* if $P(a)$ is left-computable for every $a \in \Omega$. On the other hand, we say that P is *right-computable* if $P(a)$ is right-computable for every $a \in \Omega$. However, using the condition (ii) of Definition 3 we can see that these three computable notions for a finite probability space coincide with one another, as the following proposition states.

Proposition 5. *Let $P \in \mathbb{P}(\Omega)$. The following conditions are equivalent to one another.*

- (i) P is computable.
- (ii) P is left-computable.
- (iii) P is right-computable. □

4 Martin-Löf P -randomness

In order to provide an operational characterization of the notion of probability we use a generalization of Martin-Löf randomness over Bernoulli measure.

Let $P \in \mathbb{P}(\Omega)$. For each $\sigma \in \Omega^*$, we use $P(\sigma)$ to denote $P(\sigma_1)P(\sigma_2)\dots P(\sigma_n)$ where $\sigma = \sigma_1\sigma_2\dots\sigma_n$ with $\sigma_i \in \Omega$. For each subset S of Ω^* , we use $P(S)$ to denote

$$\sum_{\sigma \in S} P(\sigma).$$

Consider a function $r: \Omega^* \rightarrow [0, 1]$ such that $r(\sigma) = P(\sigma)$ for every $\sigma \in \Omega^*$. It is then easy to see that the function r is a probability measure representation over Ω . The probability measure μ_r induced by r is called a *Bernoulli measure on Ω^∞* , denoted λ_P . The Bernoulli measure λ_P on Ω^∞ has the following property: For every $\sigma \in \Omega^*$,

$$\lambda_P([\sigma]^\prec) = P(\sigma), \tag{3}$$

which results from (2).

Martin-Löf randomness with respect to Bernoulli measure, which is called *Martin-Löf P -randomness* in this paper, is defined as follows. This notion was, in essence, introduced by Martin-Löf [14], as well as the notion of Martin-Löf randomness which we have described in Definition 1.

Definition 6 (Martin-Löf P -randomness, Martin-Löf [14]). *Let $P \in \mathbb{P}(\Omega)$.*

- (i) *A subset \mathcal{C} of $\mathbb{N}^+ \times \Omega^*$ is called a Martin-Löf P -test if \mathcal{C} is an r.e. set such that for every $n \in \mathbb{N}^+$ it holds that*

$$\lambda_P([\mathcal{C}_n]^\prec) < 2^{-n}$$

where $\mathcal{C}_n := \{ \sigma \mid (n, \sigma) \in \mathcal{C} \}$.

- (ii) *For any $\alpha \in \Omega^\infty$ and Martin-Löf P -test \mathcal{C} , we say that α passes \mathcal{C} if there exists $n \in \mathbb{N}^+$ such that $\alpha \notin [\mathcal{C}_n]^\prec$.*
- (iii) *For any $\alpha \in \Omega^\infty$, we say that α is Martin-Löf P -random if for every Martin-Löf P -test \mathcal{C} it holds that α passes \mathcal{C} . □*

Note that we do not require P to be computable in Definition 6. Thus, Bernoulli measure λ_P itself is not necessarily computable in Definition 6. Here, we say that Bernoulli measure λ_P is *computable* if there exists a computable function $g: \mathbb{N} \times \Omega^* \rightarrow \mathbb{Q}$ such that $|\lambda_P([\sigma]^\prec) - g(k, \sigma)| < 2^{-k}$ for all $k \in \mathbb{N}$ and $\sigma \in \Omega^*$. Note also that in Definition 6 we do not require that $P(a) > 0$ for all $a \in \Omega$. Therefore, $P(a_0)$ may be 0 for some $a_0 \in \Omega$. In the case where $\Omega = \{0, 1\}$ and P satisfies that $P(0) = P(1) = 1/2$, the Martin-Löf P -randomness results in the Martin-Löf randomness in Definition 1.

Since there are only countably infinitely many algorithms and every Martin-Löf P -test induces an effective null set, as we saw in Section 2.4 in the case of a Martin-Löf test, it is easy to show the following theorem.

Theorem 7. $\lambda_P(\text{ML}_P) = 1$ for every $P \in \mathbb{P}(\Omega)$, where ML_P is the set of all Martin-Löf P -random sequences over Ω . \square

5 Ensemble

In this section we present an operational characterization of the notion of probability for a finite probability space, and consider its validity. We propose to regard a Martin-Löf P -random sequence of sample points as an *operational characterization of the notion of probability* for a finite probability space P on Ω . Namely, we propose to identify a Martin-Löf P -random sequence of sample points with the *substance* of the notion of probability for a finite probability space P . Thus, since the notion of Martin-Löf P -random sequence plays a central role in our framework, in particular we call it an *ensemble*, as in Definition 8, instead of collective for distinction. The name “ensemble” comes from physics, in particular, from quantum mechanics and statistical mechanics.¹

Definition 8 (Ensemble). *Let $P \in \mathbb{P}(\Omega)$. A Martin-Löf P -random infinite sequence over Ω is called an ensemble for the finite probability space P on Ω .* \square

Let $P \in \mathbb{P}(\Omega)$. Consider an infinite sequence $\alpha \in \Omega^\infty$ of outcomes which is being generated by infinitely repeated trials *described by* the finite probability space P . The operational characterization of the notion of probability for the finite probability space P is thought to be completed if the property which the infinite sequence α has to satisfy is determined. We thus propose the following thesis.

Thesis 1. *Let $P \in \mathbb{P}(\Omega)$. An infinite sequence of outcomes in Ω which is being generated by infinitely repeated trials described by the finite probability space P on Ω is an ensemble for P .* \square

Let us check the validity of Thesis 1. First of all, what is “probability”? It would seem very difficult to answer this question *completely* and *sufficiently*. However, we may enumerate the *necessary* conditions which the notion of probability is considered to have to satisfy *according to our intuitive understanding of the notion of probability*. In the subsequent subsections, we check that the notion of ensemble satisfies these necessary conditions.

¹The notion of ensemble plays a fundamental role in quantum mechanics and statistical mechanics. However, the notion is very vague in physics from a mathematical point of view. We propose to regard a Martin-Löf P -random sequence of quantum states as a formal definition of the notion of ensemble in quantum mechanics and statistical mechanics [21, 22, 23, 24, 25, 26, 27].

5.1 Elementary event with probability one

Let $P \in \mathbb{P}(\Omega)$, and let us consider an infinite sequence $\alpha \in \Omega^\infty$ of outcomes which is being generated by infinitely repeated trials described by the finite probability space P on Ω . The first necessary condition which the notion of probability for the finite probability space P is considered to have to satisfy is the condition that *an elementary event with probability one always occurs in the infinite sequence α* , i.e., the condition that for every $a \in \Omega$ if $P(a) = 1$ then α is of the form $\alpha = aaaaaa \dots$. This intuition that *an elementary event with probability one occurs certainly* is particularly supported by the notion of probability in *quantum mechanics*, as we will see in what follows.

First, we recall some of the central postulates of quantum mechanics. For simplicity, we here consider the postulates of quantum mechanics for a *finite-dimensional* quantum system, i.e., a quantum system whose state space is a finite-dimensional Hilbert space. See e.g. Nielsen and Chuang [15, Section 2.2] for the detail of the postulates of quantum mechanics, in particular, in the finite-dimensional case. We refer to some of the postulates from it.

The first postulate of quantum mechanics is about *state space* and *state vector*.

Postulate 1 (State space and state vector). *Associated to any isolated physical system is a complex vector space with inner product (i.e., Hilbert space) known as the state space of the system. The system is completely described by its state vector, which is a unit vector in the system's state space.* \square

The second postulate of quantum mechanics is about measurements on quantum systems. This is the so-called *Born rule*, i.e., *the probability interpretation of the wave function*.

Postulate 2 (The Born rule). *Quantum measurements are described by an observable, M , a Hermitian matrix on the state space of the system being measured. The observable has a spectral decomposition*

$$M = \sum_m m E_m,$$

where E_m is the projector onto the eigenspace of M with eigenvalue m . The possible outcomes of the measurement correspond to the eigenvalues, m , of the observable. If the state of the quantum system is Ψ immediately before the measurement then the probability that result m occurs is given by $(\Psi, E_m \Psi)$, where (\cdot, \cdot) denotes the inner-product defined on the state space of the system². \square

Postulate 2 describes the effects of measurements on quantum systems using the notion of *probability*, whereas it does not mention the *operational definition* of the notion of probability. On the other hand, there is a postulate about quantum measurements with no reference to the notion of probability. This is given in Dirac [8, Section 10], and describes a special case of quantum measurements which are performed upon a quantum system in an *eigenstate* of an observable, i.e., a state represented by an eigenvector of an observable.

Postulate 3 (Dirac [8]). *If the dynamical system is in an eigenstate of a real dynamical variable ξ , belonging to the eigenvalue ξ' , then a measurement of ξ will certainly gives as result the number ξ' .* \square

²To be precise, Postulate 2 is just a part of the Born rule. The original Born rule includes a statement about the post-measurement state, i.e., the statement that immediately after the measurement where result m has occurred, the state of the system is given by $E_m \Psi / \sqrt{(\Psi, E_m \Psi)}$.

Here, the “dynamical system” means quantum system, and the “real dynamical variable” means observable.

Based on Postulates 1, 2, and 3 above, we can show that an elementary event *with probability one* occurs certainly in quantum mechanics. To see this, let us consider a quantum system with finite-dimensional state space, and a measurement described by an observable M performed upon the system. Suppose that the probability of getting result m_0 is one in the measurement performed upon the system in a state represented by a state vector Ψ . Let

$$M = \sum_m m E_m$$

be a spectral decomposition of the observable M , where E_m is the projector onto the eigenspace of M with eigenvalue m . Then, it follows from Postulate 2 that $(\Psi, E_{m_0} \Psi) = 1$. This implies that Ψ is an eigenvector of M belonging to the eigenvalue m_0 , since Ψ is a unit vector. Therefore, we have that immediately before the measurement, the quantum system is in an eigenstate of the observable M , belonging to the eigenvalue m_0 . It follows from Postulate 3 that the measurement of M will *certainly* give as result the number m_0 . Hence, it turns out that *an elementary event with probability one occurs certainly in quantum mechanics*.

Theorem 9 below states that an elementary event with probability one always occurs in an ensemble, and thus shows that the notion of ensemble coincides with our intuition about the notion of probability, in particular, in quantum mechanics.

Theorem 9. *Let $P \in \mathbb{P}(\Omega)$, and let $a \in \Omega$. Suppose that α is an ensemble for the finite probability space P and $P(a) = 1$. Then α consists only of a , i.e., $\alpha = aaaaaa \dots$. \square*

To show Theorem 9, we first show Theorem 10 below. The result (i) of Theorem 10 states that an elementary event with probability zero never occurs in an ensemble, from which Theorem 9 follows immediately. The result (i) was, in essence, pointed out by Martin-Löf [14].

Theorem 10. *Let $P \in \mathbb{P}(\Omega)$.*

- (i) *Let $a \in \Omega$. Suppose that α is an ensemble for the finite probability space P and $P(a) = 0$. Then α does not contain a .*
- (ii) *Actually, there exists a single Martin-Löf P -test $\mathcal{C} \subset \mathbb{N}^+ \times \Omega^*$ such that, for every $\alpha \in \Omega^\infty$, if α passes \mathcal{C} then α does not contain any element of $P^{-1}(\{0\})$.*

Proof. It is sufficient to prove (ii) of Theorem 10. For that purpose, we first define S as $\Omega^* \setminus (\Omega \setminus P^{-1}(\{0\}))^*$, and then define \mathcal{C} as the set $\{(n, \sigma) \mid n \in \mathbb{N}^+ \text{ \& } \sigma \in S\}$. Since $P(\sigma) = 0$ for every $\sigma \in S$, we have $\lambda_P([C_n]^\prec) \leq P(C_n) = P(S) = 0$ for each $n \in \mathbb{N}^+$, and \mathcal{C} is r.e., obviously. Hence, \mathcal{C} is Martin-Löf P -test.

Let $\alpha \in \Omega^\infty$. Suppose that α passes \mathcal{C} . Assume contrarily that α contains some element a_0 of $P^{-1}(\{0\})$. Then there exists a prefix σ_0 of α which contains a_0 . It follows that $\sigma_0 \in S$, and therefore $\alpha \in [C_n]^\prec$ for all $n \in \mathbb{N}^+$. Hence, we have a contradiction, and the proof is completed. \square

5.2 The law of large numbers

Let $P \in \mathbb{P}(\Omega)$, and let us consider an infinite sequence $\alpha \in \Omega^\infty$ of outcomes which is being generated by infinitely repeated trials described by the finite probability space P on Ω . The second necessary

condition which the notion of probability for the finite probability space P is considered to have to satisfy is the condition that *the law of large numbers holds for α* . Theorem 11 below confirms that this certainly holds. Note here that we have to prove that the law of large numbers holds for α even in the case where P is *not computable*. This is because a finite probability space is not computable, in general. However, we can certainly prove it, as shown in Theorem 11.

Theorem 11 (The law of large numbers). *Let $P \in \mathbb{P}(\Omega)$. For every $\alpha \in \Omega^\infty$, if α is an ensemble for P then for every $a \in \Omega$ it holds that*

$$\lim_{n \rightarrow \infty} \frac{N_a(\alpha \upharpoonright_n)}{n} = P(a),$$

where $N_a(\sigma)$ denotes the number of the occurrences of a in σ for every $a \in \Omega$ and $\sigma \in \Omega^*$. \square

In order to prove Theorem 11, we need the following theorem, Chernoff bound, which is a modification of the form given in Goldreich [11, Section 1.2.2].

Theorem 12 (Chernoff bound). *Let $P \in \mathbb{P}(\{0, 1\})$. Then for each ε with $0 < \varepsilon \leq P(0)P(1)$ and each $n \in \mathbb{N}^+$, we have*

$$\lambda_P([S_n]^\prec) < 2e^{-\frac{\varepsilon^2}{2P(0)P(1)}n},$$

where S_n is the set of all $\sigma \in \{0, 1\}^n$ such that $|N_1(\sigma)/n - P(1)| > \varepsilon$. \square

In order to prove Theorem 11, we also need the following theorem.

Theorem 13. *Let $P \in \mathbb{P}(\Omega)$. Let α be an ensemble for P , and let a and b be distinct elements of Ω . Suppose that β is an infinite sequence over $\Omega \setminus \{b\}$ obtained by replacing all occurrences of b by a in α . Then β is an ensemble for Q , where $Q \in \mathbb{P}(\Omega \setminus \{b\})$ such that $Q(x) := P(a) + P(b)$ if $x = a$ and $Q(x) := P(x)$ otherwise.*

Proof. We show the contraposition. Suppose that β is not Martin-Löf Q -random. Then there exists a Martin-Löf Q -test $\mathcal{S} \subset \mathbb{N}^+ \times (\Omega \setminus \{b\})^*$ such that $\beta \in [\mathcal{S}_n]^\prec$ for every $n \in \mathbb{N}^+$. For each $\sigma \in (\Omega \setminus \{b\})^*$, let $f(\sigma)$ be the set of all $\tau \in \Omega^*$ such that τ is obtained by replacing some or none of the occurrences of a in σ , if exists, by b . Note that if σ has exactly n occurrences of a then $\#f(\sigma) = 2^n$. We then define \mathcal{T} to be a subset of $\mathbb{N}^+ \times \Omega^*$ such that $\mathcal{T}_n = \bigcup_{\sigma \in \mathcal{S}_n} f(\sigma)$ for every $n \in \mathbb{N}^+$. Since $Q(a) = P(a) + P(b)$, we have $\lambda_Q([\sigma]^\prec) = Q(\sigma) = P(f(\sigma)) = \lambda_P([f(\sigma)]^\prec)$ for each $\sigma \in (\Omega \setminus \{b\})^*$. Therefore, it is easy to see that $\lambda_P([\mathcal{T}_n]^\prec) = \lambda_Q([\mathcal{S}_n]^\prec) < 2^{-n}$ for each $n \in \mathbb{N}^+$. Since \mathcal{S} is r.e., \mathcal{T} is also r.e. Thus, \mathcal{T} is Martin-Löf P -test. On the other hand, it follows that $\alpha \in [\mathcal{T}_n]^\prec$ for every $n \in \mathbb{N}^+$. Hence, α is not Martin-Löf P -random. This completes the proof. \square

Theorem 11 is then proved as follows.

Proof of Theorem 11. Let $a \in \Omega$. In the case of $P(a) = 0$, the result follows immediately from (i) of Theorem 10. In the case of $P(a) = 1$, the result follows immediately from Theorem 9. Thus we assume that $0 < P(a) < 1$, in what follows.

We define $Q \in \mathbb{P}(\{0, 1\})$ by the condition that $Q(1) = P(a)$ and $Q(0) = 1 - P(a)$. Then $Q(0)Q(1) > 0$. Let β be the infinite binary sequence obtained from α by replacing all a by 1 and all other elements of Ω by 0 in α . Then, by using Theorem 13 repeatedly, it is easy to show that β is Martin-Löf Q -random and $N_1(\beta \upharpoonright_n) = N_a(\alpha \upharpoonright_n)$ for every $n \in \mathbb{N}^+$.

Assume contrarily that $\lim_{n \rightarrow \infty} N_a(\alpha \upharpoonright_n)/n \neq P(a)$. Then $\lim_{n \rightarrow \infty} N_1(\beta \upharpoonright_n)/n \neq Q(1)$ and therefore there exists $\varepsilon > 0$ such that $|N_1(\beta \upharpoonright_n)/n - Q(1)| > 2\varepsilon$ for infinitely many $n \in \mathbb{N}^+$. On the other hand, it follows from Theorem 12 that

$$\lambda_Q([\{\sigma \in \{0,1\}^n \mid |N_1(\sigma)/n - Q(1)| > \varepsilon\}]^\prec) < 2e^{-\frac{\varepsilon^2}{2Q(0)Q(1)}n}$$

for every $n \in \mathbb{N}^+$. Since $Q(1)$ is not necessarily computable, we choose $r_L, r_R \in \mathbb{Q}$ such that $Q(1) - 2\varepsilon < r_L < Q(1) - \varepsilon$ and $Q(1) + \varepsilon < r_R < Q(1) + 2\varepsilon$. For each $n \in \mathbb{N}^+$, let S_n be the set $\{\sigma \in \{0,1\}^n \mid N_1(\sigma)/n < r_L \text{ or } r_R < N_1(\sigma)/n\}$ and let $T_n = \bigcup_{m=n}^{\infty} S_m$. Then, for every $n \in \mathbb{N}^+$ it holds that $\beta \in [T_n]^\prec$ and

$$\lambda_Q([T_n]^\prec) \leq \sum_{m=n}^{\infty} 2e^{-cm} = \frac{2e^{-cn}}{1 - e^{-c}},$$

where c is a specific rational with $0 < c < \varepsilon^2/(2Q(0)Q(1))$. Then it is easy to show that there exists a total recursive function $f: \mathbb{N}^+ \rightarrow \mathbb{N}^+$ such that $2e^{-cf(n)}/(1 - e^{-c}) < 2^{-n}$. Thus, β is not Martin-Löf Q -random since the set $\{(n, \sigma) \mid n \in \mathbb{N}^+ \text{ \& } \sigma \in T_{f(n)}\}$ is Martin-Löf Q -test and $\beta \in [T_{f(n)}]^\prec$ for every $n \in \mathbb{N}^+$. Hence we have a contradiction, and the result follows. \square

We here remark that the notion of probability is more than the law of large numbers. To see this, let us consider a finite probability space $P \in \mathbb{P}(\{a, b\})$ such that $P(a) = 1$ and $P(b) = 0$, and consider an infinite sequence

$$\alpha = abaaaaaaaaa \dots$$

over $\{a, b\}$. Then, since $\lim_{n \rightarrow \infty} N_a(\alpha \upharpoonright_n)/n = 1 = P(a)$, the law of large numbers certainly holds for α . However, the elementary event b with probability zero has occurred once in α . This contradicts our intuition that *an elementary event with probability one always occurs*, which is our conclusion in the preceding subsection from the aspect of the notion of probability, in particular, in quantum mechanics. Thus, the example shows that the law of large numbers is insufficient to characterize the notion of probability, and *the notion of probability is more than the law of large numbers*.

The following is immediate from Theorem 11.

Corollary 14. *Let $P, Q \in \mathbb{P}(\Omega)$. If there exists $\alpha \in \Omega^\infty$ which is both an ensemble for P and an ensemble for Q , then $P = Q$.* \square

5.3 Computable shuffling

This subsection considers the third necessary condition which the notion of probability for a finite probability space is considered to have to satisfy.

Let $P \in \mathbb{P}(\Omega)$. Assume that an observer A performs an infinite reputation of trials described by the finite probability space P , and thus is generating an infinite sequence $\alpha \in \Omega^\infty$ of outcomes of the trials as

$$\alpha = a_1 a_2 a_3 a_4 a_5 a_6 a_7 a_8 \dots$$

with $a_i \in \Omega$. According to our thesis, Thesis 1, α is an ensemble for P . Consider another observer B who wants to adopt the following subsequence β of α as the outcomes of the trials:

$$\beta = a_2 a_3 a_5 a_7 a_{11} a_{13} a_{17} \dots,$$

where the observer B only takes into account the n th elements a_n in the original sequence α such that n is a prime number. According to Thesis 1, β has to be an ensemble for P , as well. However, is this true?

Consider this problem in a general setting. Assume as before that an observer A performs an infinite reputation of trials described by the finite probability space P , and thus is generating an infinite sequence $\alpha \in \Omega^\infty$ of outcomes of the trials. According to Thesis 1, α is an ensemble for P . Now, let $f: \mathbb{N}^+ \rightarrow \mathbb{N}^+$ be an injection. Consider another observer B who wants to adopt the following sequence β as the outcomes of the trials:

$$\beta = \alpha(f(1))\alpha(f(2))\alpha(f(3))\alpha(f(4))\alpha(f(5))\dots\dots$$

instead of α . According to Thesis 1, β has to be an ensemble for P , as well. However, is this true?

We can confirm this *by restricting the ability of B* , that is, by assuming that every observer can select elements from the original sequence α *only in an effective manner*. This means that the function $f: \mathbb{N}^+ \rightarrow \mathbb{N}^+$ has to be a computable function. Theorem 15 below shows this result.

In other words, Theorem 15 states that ensembles for P are *closed under computable shuffling*.

Theorem 15 (Closure property under computable shuffling). *Let $P \in \mathbb{P}(\Omega)$, and let α be an ensemble for P . Then, for every injective function $f: \mathbb{N}^+ \rightarrow \mathbb{N}^+$, if f is computable then the infinite sequence*

$$\alpha_f := \alpha(f(1))\alpha(f(2))\alpha(f(3))\alpha(f(4))\dots\dots\dots$$

is an ensemble for P .

Proof. We show the contraposition. Suppose that α_f is not Martin-Löf P -random. Then there exists a Martin-Löf P -test $\mathcal{C} \subset \mathbb{N}^+ \times \Omega^*$ such that $\alpha_f \in [\mathcal{C}_n]^\prec$ for every $n \in \mathbb{N}^+$. For each $\sigma \in \Omega^+$, let $F(\sigma)$ be the set of all $\tau \in \Omega^+$ such that

- (i) $|\tau| = \max f(\{1, 2, \dots, |\sigma|\})$, and
- (ii) for every $k = 1, 2, \dots, |\sigma|$ it holds that $\sigma(k) = \tau(f(k))$.

We then define \mathcal{D} to be a subset of $\mathbb{N}^+ \times \Omega^*$ such that $\mathcal{D}_n = \bigcup_{\sigma \in \mathcal{C}_n} F(\sigma)$ for every $n \in \mathbb{N}^+$. Note here that, for each $n \in \mathbb{N}^+$, $\lambda \notin \mathcal{C}_n$ since $\lambda_P([\mathcal{C}_n]^\prec) < 2^{-n} < 1$. Since f is an injection and $\sum_{a \in \Omega} P(a) = 1$, we have $\lambda_P([F(\sigma)]^\prec) = P(F(\sigma)) = P(\sigma) = \lambda_P([\sigma]^\prec)$ for each $\sigma \in \Omega^+$. Therefore, it is easy to see that $\lambda_P([\mathcal{D}_n]^\prec) = \lambda_P([\mathcal{C}_n]^\prec) < 2^{-n}$ for each $n \in \mathbb{N}^+$. Since \mathcal{C} is r.e., \mathcal{D} is also r.e. Thus, \mathcal{D} is Martin-Löf P -test. On the other hand, it follows that $\alpha \in [\mathcal{D}_n]^\prec$ for every n . Hence, α is not Martin-Löf P -random. This completes the proof. \square

5.4 Selection by partial computable selection functions

As the forth necessary condition which the notion of probability for a finite probability space P on Ω is considered to have to satisfy, in this subsection we consider the condition that infinite sequences over Ω of outcomes each of which is obtained by an infinite reputation of the trials described by the finite probability space P are *closed under the selection by a partial computable selection function used in the definition of von Mises-Wald-Church stochasticity*. The notion of von Mises-Wald-Church stochasticity is investigated in the theory of collectives [30, 31, 32, 33, 7].³ For

³See Downey and Hirschfeldt [9, Section 7.4] for a treatment of the mathematics of the notion of von Mises-Wald-Church stochasticity *itself* from a modern point of view.

motivating the forth necessary condition, we carry out a thought experiment in what follows, as in the preceding subsection.

Let $P \in \mathbb{P}(\Omega)$, and let us assume that an observer A performs an infinite reputation of trials described by the finite probability space P , and thus is generating an infinite sequence $\alpha \in \Omega^\infty$ of outcomes of the trials as

$$\alpha = a_1 a_2 a_3 a_4 a_5 a_6 \dots\dots$$

with $a_i \in \Omega$. According to Thesis 1, α is an ensemble for P .

Consider another observer B who wants to *refute* Thesis 1. For that purpose, the observer B adopts a subsequence $\beta = b_1 b_2 b_3 b_4 \dots\dots$ with $b_i \in \Omega$ of α in the following manner: Whenever a new outcome a_n is generated by the observer A , the observer B investigates the prefix $a_1 a_2 a_3 \dots a_n$ of α generated so far by the observer A . Then, based on the prefix, the observer B decides whether the next outcome a_{n+1} should be appended to the tail of $b_1 b_2 b_3 \dots b_k$ which have been adopted so far by B as a prefix of β . In this manner the observer B is generating the subsequence β of α . Note that the length of β may or may not be infinite.

On the other hand, the observer A is a *defender* of Thesis 1. Therefore, the observer A tries to inhibit the observer B from breaking Thesis 1. For that purpose, the observer A never generates the next outcome a_{n+1} before the observer B decides whether this a_{n+1} should be appended to the tail of $b_1 b_2 b_3 \dots b_k$. This is because if for each n the observer B knows the outcome a_{n+1} before the decision for a_{n+1} to be appended or to be ignored, then the observer B can easily generate an infinite subsequence β of α which does not satisfy Thesis 1. Thus, due to this careful behavior of the observer A , the observer B has to make the decision of the choice of the next outcome a_{n+1} , based only on the prefix $a_1 a_2 a_3 \dots a_n$ of α , without knowing the outcome a_{n+1} . Then, according to Thesis 1, β has to be an ensemble for P , as well as α . However, is this true?

We can confirm this by restricting the ability of B , that is, by assuming that the observer B can make the decision of the choice of the next outcome, *only in an effective manner* based on the prefix $a_1 a_2 a_3 \dots a_n$ of α generated so far by the observer A .

Put more mathematically, we introduce some notations. A *selection function* is a partial function $f: \Omega^* \rightarrow \{\text{YES}, \text{NO}\}$. We think of f as the decision of B whether or not to choose the next outcome $\alpha(n+1)$ based on the prefix $\alpha|_n$ of α in generating β . For any $\gamma \in \Omega^\infty$, $k \in \mathbb{N}^+$, and selection function g , let $s_g(\gamma, k)$ be the k th number $\ell \in \mathbb{N}$ such that $g(\gamma|_\ell) = \text{YES}$, i.e., the least number $m \in \mathbb{N}$ such that $\#\{\ell \leq m \mid g(\gamma|_\ell) = \text{YES}\} = k$, if such k exists.

First, consider the case where $f(\alpha|_n)$ is not defined for some $n \in \mathbb{N}$. Let m be the least number of such n . Then, this case means that the observer B does not make the decision of the choice of the next outcome $\alpha(m+1)$ based on the prefix $\alpha|_m$, and is stalled. Therefore, the length of β is finite in this case. Thus, the observer B *cannot refute* Thesis 1 in this case, since Thesis 1 only refers to the property of an *infinite* sequence of outcomes which is being generated by *infinitely* repeated trials. Hence, Thesis 1 survives in this case.

Secondly, consider the case where $f(\alpha|_n)$ is defined for all $n \in \mathbb{N}$ and $\{n \in \mathbb{N} \mid f(\alpha|_n) = \text{YES}\}$ is a finite set. In this case, the length of β is also finite. Thus, the observer B does not refute Thesis 1, and therefore Thesis 1 survives also in this case.

Finally, consider the remaining case, where $f(\alpha|_n)$ is defined for all $n \in \mathbb{N}$ and the set $\{n \in \mathbb{N} \mid f(\alpha|_n) = \text{YES}\}$ is infinite. Then, $s_f(\alpha, k)$ is defined and $\beta(k) = \alpha(s_f(\alpha, k) + 1)$ for all $k \in \mathbb{N}^+$. Hence, β is an infinite sequence over Ω , and thus Thesis 1 can be applied to β in this case. Therefore, according to Thesis 1, β has to be an ensemble for P , as well as α . However, is this true? Actually, we can confirm this *by restricting the ability of B* , that is, *by assuming that f has to be a partial*

computable selection function. Here, a selection function $g: \Omega^* \rightarrow \{\text{YES}, \text{NO}\}$ is called a *partial computable selection function* if $g: \Omega^* \rightarrow \{\text{YES}, \text{NO}\}$ is a partial computable function. Theorem 16 below shows this result. It states that ensembles for an arbitrary finite probability space are *closed under the selection by a partial computable selection function*. Hence, Thesis 1 survives in this case as well.

In this way, based on Theorem 16, we confirm that the forth condition certainly holds for ensembles for an arbitrary finite probability space.

Theorem 16 (Closure property under the selection by a partial computable selection function). *Let $P \in \mathbb{P}(\Omega)$, and let α be an ensemble for P . Let f be a partial computable selection function. Suppose that $f(\alpha \upharpoonright_k)$ is defined for all $k \in \mathbb{N}$ and $\{k \in \mathbb{N} \mid f(\alpha \upharpoonright_k) = \text{YES}\}$ is an infinite set. Then an infinite sequence β such that $\beta(k) = \alpha(s_f(\alpha, k) + 1)$ for all $k \in \mathbb{N}^+$ is an ensemble for P .*

Proof. We show the contraposition. Suppose that β is not Martin-Löf P -random. Then there exists a Martin-Löf P -test $\mathcal{C} \subset \mathbb{N}^+ \times \Omega^*$ such that $\beta \in [\mathcal{C}_n]^\prec$ for every $n \in \mathbb{N}^+$. For each $\sigma, \tau \in \Omega^+$, we say that σ is *selected by f from τ* if $f(\tau \upharpoonright_k)$ is defined for all $k = 0, 1, \dots, |\tau| - 1$ and there exists a strictly increasing function $h: \{1, \dots, |\sigma|\} \rightarrow \mathbb{N}$ such that

- (i) $\{k \in \{1, \dots, |\tau|\} \mid f(\tau \upharpoonright_{k-1}) = \text{YES}\} = h(\{1, \dots, |\sigma|\})$,
- (ii) $h(|\sigma|) = |\tau|$, and
- (iii) $\tau(h(k)) = \sigma(k)$ for all $k = 1, \dots, |\sigma|$.

For each $\sigma \in \Omega^+$, let $F(\sigma)$ be the set of all $\tau \in \Omega^*$ such that σ is selected by f from τ . We also set $F(\lambda) := \{\lambda\}$. It is then easy to see that $F(\sigma)$ is a prefix-free set for every $\sigma \in \Omega^*$.

We show that

$$\lambda_P([F(\sigma)]^\prec) \leq \lambda_P([\sigma]^\prec) \quad (4)$$

for all $\sigma \in \Omega^*$ by the induction on $|\sigma|$. First, the inequality (4) holds for the case of $|\sigma| = 0$, obviously. For an arbitrary $n \in \mathbb{N}$, assume that (4) holds for all $\sigma \in \Omega^n$. Let $\sigma \in \Omega^{n+1}$. We then denote the prefix of σ of length n by ρ , and denote $\sigma(|\sigma|)$ by a . Therefore $\sigma = \rho a$. Note that

$$G(\tau) := \{v \in \Omega^* \mid \tau v a \in F(\sigma)\}$$

is a prefix-free set for each $\tau \in \Omega^*$. Therefore, we have

$$\begin{aligned} \lambda_P([F(\sigma)]^\prec) &= \sum_{\tau \in F(\sigma)} \lambda_P([\tau]^\prec) = \sum_{\tau \in F(\rho)} \sum_{v \in G(\tau)} \lambda_P([\tau v a]^\prec) \\ &= \sum_{\tau \in F(\rho)} \sum_{v \in G(\tau)} \lambda_P([\tau]^\prec) \lambda_P([v]^\prec) P(a) \\ &\leq \sum_{\tau \in F(\rho)} \lambda_P([\tau]^\prec) P(a) = \lambda_P([F(\rho)]^\prec) P(a) \\ &\leq \lambda_P([\rho]^\prec) P(a) = \lambda_P([\sigma]^\prec), \end{aligned}$$

where the second inequality follows from the assumption.

We then define $\mathcal{D} \subset \mathbb{N}^+ \times \Omega^*$ by the condition that $\mathcal{D}_n = \bigcup_{\sigma \in \mathcal{C}_n} F(\sigma)$ for every $n \in \mathbb{N}^+$. It follows from (4) that

$$\lambda_P([\mathcal{D}_n]^\prec) \leq \sum_{\sigma \in \mathcal{C}_n} \lambda_P([F(\sigma)]^\prec) \leq \sum_{\sigma \in \mathcal{C}_n} \lambda_P([\sigma]^\prec) = \lambda_P([\mathcal{C}_n]^\prec) < 2^{-n}$$

for each $n \in \mathbb{N}^+$. Thus, since \mathcal{D} is r.e., we see that \mathcal{D} is Martin-Löf P -test. On the other hand, it is easy to see that $\alpha \in [\mathcal{D}_n]^\prec$ for every $n \in \mathbb{N}^+$. Therefore, α is not Martin-Löf P -random. This completes the proof. \square

Theorem 15 and Theorem 16 show that certain closure properties hold for ensembles for an arbitrary finite probability space. In the subsequent sections, we will see that various strong closure properties of another type hold for the ensembles.

6 Conditional probability and the independence between two events

In this section we operationally characterize the notions of *conditional probability* and the *independence between two events* on a finite probability space, in terms of ensembles.

Let $P \in \mathbb{P}(\Omega)$, and let $A \subset \Omega$ be an event on the finite probability space P . For each ensemble α for P , we use $C_A(\alpha)$ to denote the infinite binary sequence such that, for every $n \in \mathbb{N}^+$, its n th element $(C_A(\alpha))(n)$ is 1 if $\alpha(n) \in A$ and 0 otherwise. The pair (P, A) induces a finite probability space $C(P, A) \in \mathbb{P}(\{0, 1\})$ such that $(C(P, A))(1) = P(A)$ and $(C(P, A))(0) = 1 - P(A)$. Note that the notions of $C_A(\alpha)$ and $C(P, A)$ in our theory together correspond to the notion of *mixing* in the theory of collectives by von Mises [31]. We can then show the following theorem.

Theorem 17. *Let $P \in \mathbb{P}(\Omega)$, and let $A \subset \Omega$. Suppose that α is an ensemble for the finite probability space P . Then $C_A(\alpha)$ is an ensemble for the finite probability space $C(P, A)$.*

Proof. We show the contraposition. Suppose that $C_A(\alpha)$ is not Martin-Löf $C(P, A)$ -random. Then there exists a Martin-Löf $C(P, A)$ -test $\mathcal{S} \subset \mathbb{N}^+ \times \{0, 1\}^*$ such that $C_A(\alpha) \in [\mathcal{S}_n]^\prec$ for every $n \in \mathbb{N}^+$. For each $\sigma \in \{0, 1\}^*$, let $f(\sigma)$ be the set of all $\tau \in \Omega^*$ such that τ is obtained by replacing each occurrence of 1 in σ , if exists, by some of element of A and by replacing each occurrence of 0 in σ , if exists, by some of element of $\Omega \setminus A$. For example, if $\Omega = \{x, y, z\}$ and $A = \{x, y\}$ then $f(011) = \{zxx, zxy, zyx, zyy\}$. We then define \mathcal{T} to be a subset of $\mathbb{N}^+ \times \Omega^*$ such that $\mathcal{T}_n = \bigcup_{\sigma \in \mathcal{S}_n} f(\sigma)$ for every $n \in \mathbb{N}^+$. Since $(C(P, A))(1) = \sum_{a \in A} P(a)$ and $(C(P, A))(0) = \sum_{a \in \Omega \setminus A} P(a)$, we have $\lambda_{C(P, A)}([\sigma]^\prec) = (C(P, A))(\sigma) = P(f(\sigma)) = \lambda_P([f(\sigma)]^\prec)$ for each $\sigma \in \{0, 1\}^*$. Therefore, it is easy to see that $\lambda_P([\mathcal{T}_n]^\prec) = \lambda_{C(P, A)}([\mathcal{S}_n]^\prec) < 2^{-n}$ for each $n \in \mathbb{N}^+$. Since \mathcal{S} is r.e., \mathcal{T} is also r.e. Thus, \mathcal{T} is Martin-Löf P -test. On the other hand, it follows that $\alpha \in [\mathcal{T}_n]^\prec$ for every $n \in \mathbb{N}^+$. Hence, α is not Martin-Löf P -random. This completes the proof. \square

We show that the notion of conditional probability in a finite probability space can be represented by an ensemble in a natural manner. For that purpose, first we recall the notion of conditional probability in a finite probability space.

Let $P \in \mathbb{P}(\Omega)$, and let $B \subset \Omega$ be an event on the finite probability space P . Suppose that $P(B) > 0$. Then, for each event $A \subset \Omega$, the *conditional probability of A given B* , denoted $P(A|B)$, is defined as $P(A \cap B)/P(B)$. This notion defines a finite probability space $P_B \in \mathbb{P}(B)$ such that $P_B(a) = P(\{a\}|B)$ for every $a \in B$.

When an infinite sequence $\alpha \in \Omega^\infty$ contains infinitely many elements from B , $\text{Filtered}_B(\alpha)$ is defined as an infinite sequence in B^∞ obtained from α by eliminating all elements of $\Omega \setminus B$ occurring in α . If α is an ensemble for the finite probability space P and $P(B) > 0$, then α contains infinitely many elements from B due to Theorem 11. Therefore, $\text{Filtered}_B(\alpha)$ is properly defined in this case. Note that the notion of $\text{Filtered}_B(\alpha)$ in our theory corresponds to the notion of *partition* in the theory of collectives by von Mises [31].

We can then show Theorem 18 below, which states that ensembles are *closed under conditioning*.

Theorem 18 (Closure property under conditioning). *Let $P \in \mathbb{P}(\Omega)$, and let $B \subset \Omega$ be an event on the finite probability space P with $P(B) > 0$. For every ensemble α for P , it holds that $\text{Filtered}_B(\alpha)$ is an ensemble for the finite probability space P_B .*

Proof. In the case of $B = \Omega$, we have $P_B = P$ and $\text{Filtered}_B(\alpha) = \alpha$. Therefore the result is obvious. Thus, in what follows, we assume that B is a proper subset of Ω .

First, we choose any one $a \in \Omega \setminus B$ and define a finite probability space $Q \in \mathbb{P}(B \cup \{a\})$ by the condition that $Q(x) = \sum_{y \in \Omega \setminus B} P(y)$ if $x = a$ and $Q(x) = P(x)$ otherwise. Note here that

$$1 - Q(a) = P(B), \quad (5)$$

and therefore

$$Q(a) < 1. \quad (6)$$

Let β be the infinite sequence over $B \cup \{a\}$ obtained by replacing all occurrences of elements of $\Omega \setminus B$ in α by a . It follows from Theorem 13 that β is Martin-Löf Q -random. Thus, it is sufficient to show that if $\text{Filtered}_B(\alpha)$ is not Martin-Löf P_B -random then β is not Martin-Löf Q -random.

Thus, let us assume that $\text{Filtered}_B(\alpha)$ is not Martin-Löf P_B -random. Then there exists a Martin-Löf P_B -test $\mathcal{C} \subset B \times \mathbb{N}^+$ such that

$$\text{Filtered}_B(\alpha) \in [\mathcal{C}_n]^\prec \quad (7)$$

for every $n \in \mathbb{N}^+$. For each $\sigma \in B^+$, let $F(\sigma)$ be the set of all finite strings over $B \cup \{a\}$ of the form $a^{k_1}\sigma_1 a^{k_2}\sigma_2 \dots \sigma_{L-1} a^{k_L}\sigma_L$ for some $k_1, k_2, \dots, k_L \in \mathbb{N}$, where $\sigma = \sigma_1\sigma_2 \dots \sigma_L$ with $\sigma_i \in B$. We then see that, for each $\sigma \in B^+$,

$$\begin{aligned} \lambda_Q([F(\sigma)]^\prec) &= \sum_{k_1, k_2, \dots, k_L=0}^{\infty} \lambda_Q\left(\left[a^{k_1}\sigma_1 a^{k_2}\sigma_2 \dots \sigma_{L-1} a^{k_L}\sigma_L\right]^\prec\right) \\ &= \sum_{k_1, k_2, \dots, k_L=0}^{\infty} \lambda_Q([\sigma]^\prec) Q(a)^{k_1} Q(a)^{k_2} \dots Q(a)^{k_L} \\ &= \lambda_Q([\sigma]^\prec) \left(\sum_{k=0}^{\infty} Q(a)^k\right)^L \\ &= \lambda_Q([\sigma]^\prec) \frac{1}{(1 - Q(a))^L} \\ &= \lambda_Q([\sigma]^\prec) \frac{1}{P(B)^L} \\ &= \lambda_{P_B}([\sigma]^\prec), \end{aligned}$$

where we use (6) and (5) in the forth and fifth equalities, respectively. We then define \mathcal{D} to be $\{(n, F(\sigma)) \mid n \in \mathbb{N}^+ \text{ \& } \sigma \in \mathcal{C}_n\}$. It follows that $\lambda_Q([\mathcal{D}_n]^\prec) = \lambda_{P_B}([\mathcal{C}_n]^\prec) < 2^{-n}$ for each $n \in \mathbb{N}^+$. It is easy to see that \mathcal{D} is r.e., and therefore \mathcal{D} is Martin-Löf Q -test. On the other hand, since $\text{Filtered}_B(\alpha)$ is the infinite sequence over B obtained from β by eliminating all occurrences of the symbol a in β , it follows from (7) that $\beta \in [\mathcal{D}_n]^\prec$ for every $n \in \mathbb{N}^+$. Hence, β is not Martin-Löf Q -random, and the proof is completed. \square

As an application of Theorem 18, we consider the Von Neumann extractor as follows.

Example 19 (Von Neumann extractor). *In the terminology of the conventional probability theory, consider a Bernoulli sequence. The Von Neumann extractor takes successive pairs of consecutive bits from the Bernoulli sequence. If the two bits matches, no output is generated. If the bits differ, the value of the first bit is output. The Von Neumann extractor can be shown to produce a uniform binary output. For the detail, see e.g., [35].*

In our framework, the Von Neumann extractor operates as follows: Let $P \in \mathbb{P}(\{0, 1\})$ and let α be an ensemble for P . Then α can be regarded as an ensemble for a finite probability space $Q \in \mathbb{P}(\{00, 01, 10, 11\})$ where $Q(ab) = P(a)P(b)$ for every $a, b \in \{0, 1\}$. Consider the event $B = \{01, 10\}$ on Q . It follows from Theorem 18 that $\text{Filtered}_B(\alpha)$ is an ensemble for $Q_B \in \mathbb{P}(\{01, 10\})$ with $Q_B(01) = Q_B(10) = 1/2$. Namely, α is a Martin-Löf random infinite sequence over the alphabet $\{01, 10\}$ instead of $\{0, 1\}$. Hence, a random individual infinite sequence is certainly extracted by the Von Neumann extractor in our framework. \square

Let $P \in \mathbb{P}(\Omega)$. For any events $A, B \subset \Omega$ on the finite probability space P , we say that A and B are *independent on P* if $P(A \cap B) = P(A)P(B)$. In the case of $P(B) > 0$, it holds that A and B are independent on P if and only if $P(A|B) = P(A)$.

Theorem 20 below gives operational characterizations of the notion of the independence between two events in terms of ensembles. For any $\alpha, \beta \in \Omega^\infty$, we say that α and β are *equivalent* if there exists $P \in \mathbb{P}(\Omega)$ such that α and β are both an ensemble for P .

Theorem 20. *Let $P \in \mathbb{P}(\Omega)$, and let $A, B \subset \Omega$ be events on the finite probability space P . Suppose that $P(B) > 0$. Then the following conditions are equivalent to one another.*

- (i) *The events A and B are independent on P .*
- (ii) *For every ensemble α for the finite probability space P , it holds that $C_A(\alpha)$ is equivalent to $C_{A \cap B}(\text{Filtered}_B(\alpha))$.*
- (iii) *There exists an ensemble α for the finite probability space P such that $C_A(\alpha)$ is equivalent to $C_{A \cap B}(\text{Filtered}_B(\alpha))$.*

Proof. Suppose that α is an arbitrary ensemble for the finite probability space P . Then, on the one hand, it follows from Theorem 17 that $C_A(\alpha)$ is Martin-Löf $C(P, A)$ -random. On the other hand, it follows from $P(B) > 0$ and Theorem 18 that $\text{Filtered}_B(\alpha)$ is an ensemble for the finite probability space P_B . Therefore, by Theorem 17, we see that $C_{A \cap B}(\text{Filtered}_B(\alpha))$ is Martin-Löf $C(P_B, A \cap B)$ -random.

Assume that the condition (i) holds. Then $P_B(A \cap B) = P(A)$. It follows that $C(P_B, A \cap B) = C(P, A)$. Therefore, for an arbitrary ensemble α for the finite probability space P , we see that $C_A(\alpha)$ and $C_{A \cap B}(\text{Filtered}_B(\alpha))$ are equivalent. Thus, we have the implication (i) \Rightarrow (ii).

Since there exists an ensemble α for the finite probability space P by Theorem 7, the implication (ii) \Rightarrow (iii) is obvious.

Finally, the implication (iii) \Rightarrow (i) is shown as follows. Assume that the condition (iii) holds. Then $C_A(\alpha)$ and $C_{A \cap B}(\text{Filtered}_B(\alpha))$ are Martin-Löf Q -random for some ensemble α for the finite probability space P and some $Q \in \mathbb{P}(\{0, 1\})$. It follows from the consideration at the beginning of this proof that $C_A(\alpha)$ is Martin-Löf $C(P, A)$ -random, and $C_{A \cap B}(\text{Filtered}_B(\alpha))$ is Martin-Löf $C(P_B, A \cap B)$ -random. Using Corollary 14 we see that $C(P, A) = Q = C(P_B, A \cap B)$, and therefore $P(A) = P_B(A \cap B)$. This completes the proof. \square

7 The independence of an arbitrary number of events/random variables

In this section we operationally characterize the notion of the *independence of an arbitrary number of events/random variables* on a finite probability space in terms of ensembles.

First, we consider the operational characterizations of the notion of the independence of an arbitrary number of random variables, in terms of ensembles. Let P be an arbitrary finite probability space on Ω . A *random variable* on Ω is a function $X: \Omega \rightarrow \Omega'$ where Ω' is an alphabet. Let $X_1: \Omega \rightarrow \Omega_1, \dots, X_n: \Omega \rightarrow \Omega_n$ be random variables on Ω . For any predicate $F(v_1, \dots, v_n)$ with variables v_1, \dots, v_n , we use $F(X_1, \dots, X_n)$ to denote the event

$$\{a \in \Omega \mid F(X_1(a), \dots, X_n(a))\}$$

on P . We say that the random variables X_1, \dots, X_n are *independent on P* if for every $x_1 \in \Omega_1, \dots, x_n \in \Omega_n$ it holds that

$$P(X_1 = x_1 \ \& \ \dots \ \& \ X_n = x_n) = P(X_1 = x_1) \cdots P(X_n = x_n).$$

We use $X_1 \times \cdots \times X_n$ to denote a random variable $Y: \Omega \rightarrow \Omega_1 \times \cdots \times \Omega_n$ on Ω such that

$$Y(a) = (X_1(a), \dots, X_n(a))$$

for every $a \in \Omega$.

For any random variable $X: \Omega \rightarrow \Omega'$ on Ω , we use $X(P)$ to denote a finite probability space $P' \in \mathbb{P}(\Omega')$ such that $P'(x) = P(X = x)$ for every $x \in \Omega'$. Let $\Omega_1, \dots, \Omega_n$ be alphabets. For any $P_1 \in \mathbb{P}(\Omega_1), \dots, P_n \in \mathbb{P}(\Omega_n)$, we use $P_1 \times \cdots \times P_n$ to denote a finite probability space $Q \in \mathbb{P}(\Omega_1 \times \cdots \times \Omega_n)$ such that $Q(a_1, \dots, a_n) = P_1(a_1) \cdots P_n(a_n)$ for every $a_1 \in \Omega_1, \dots, a_n \in \Omega_n$. Then the notion of the independence of random variables can be rephrased as follows.

Proposition 21. *Let $P \in \mathbb{P}(\Omega)$, and let $X_1: \Omega \rightarrow \Omega_1, \dots, X_n: \Omega \rightarrow \Omega_n$ be random variables on Ω . Then the random variables X_1, \dots, X_n are independent on P if and only if*

$$(X_1 \times \cdots \times X_n)(P) = X_1(P) \times \cdots \times X_n(P).$$

Proof. Let $x_1 \in \Omega_1, \dots, x_n \in \Omega_n$. On the one hand, we have

$$\begin{aligned} ((X_1 \times \cdots \times X_n)(P))(x_1, \dots, x_n) &= P((X_1 \times \cdots \times X_n) = (x_1, \dots, x_n)) \\ &= P(X_1 = x_1 \ \& \ \dots \ \& \ X_n = x_n). \end{aligned}$$

On the other hand, we have

$$\begin{aligned}(X_1(P) \times \cdots \times X_n(P))(x_1, \dots, x_n) &= (X_1(P))(x_1) \cdots (X_n(P))(x_n) \\ &= P(X_1 = x_1) \cdots P(X_n = x_n).\end{aligned}$$

Thus, the result follows. \square

Let $X: \Omega \rightarrow \Omega'$ be a random variable on Ω . For any $\alpha \in \Omega^\infty$, we use $X(\alpha)$ to denote an infinite sequence β over Ω' such that $\beta(k) = X(\alpha(k))$ for every $k \in \mathbb{N}^+$. We can then show the following theorem, which states that ensembles are *closed under the mapping by a random variable*.

Theorem 22 (Closure property under the mapping by a random variable). *Let $P \in \mathbb{P}(\Omega)$, and let $X: \Omega \rightarrow \Omega'$ be a random variable on Ω . If α is an ensemble for P then $X(\alpha)$ is an ensemble for $X(P)$.*

Proof. We show the contraposition. Suppose that $X(\alpha)$ is not Martin-Löf $X(P)$ -random. Then there exists a Martin-Löf $X(P)$ -test $\mathcal{S} \subset \mathbb{N}^+ \times (\Omega')^*$ such that $X(\alpha) \in [\mathcal{S}_n]^\prec$ for every $n \in \mathbb{N}^+$. For each $\sigma \in (\Omega')^+$, let $f(\sigma)$ be the set of all $\tau \in \Omega^+$ such that (i) $|\tau| = |\sigma|$ and (ii) $X(\tau(k)) = \sigma(k)$ for every $k = 1, 2, \dots, |\sigma|$. We then define \mathcal{T} to be a subset of $\mathbb{N}^+ \times \Omega^*$ such that $\mathcal{T}_n = \bigcup_{\sigma \in \mathcal{S}_n} f(\sigma)$ for every $n \in \mathbb{N}^+$. Since $(X(P))(x) = \sum_{a \in X^{-1}(x)} P(a)$ for every $x \in \Omega'$, we have $\lambda_{X(P)}([\sigma]^\prec) = (X(P))(\sigma) = P(f(\sigma)) = \lambda_P([f(\sigma)]^\prec)$ for each $\sigma \in (\Omega')^+$. Therefore, it is easy to see that $\lambda_P([\mathcal{T}_n]^\prec) = \lambda_{X(P)}([\mathcal{S}_n]^\prec) < 2^{-n}$ for each $n \in \mathbb{N}^+$. Since \mathcal{S} is r.e., \mathcal{T} is also r.e. Thus, \mathcal{T} is Martin-Löf P -test. On the other hand, it follows that $\alpha \in [\mathcal{T}_n]^\prec$ for every $n \in \mathbb{N}^+$. Hence, α is not Martin-Löf P -random. This completes the proof. \square

We introduce the notion of the *independence* of ensembles as follows. Let $\Omega_1, \dots, \Omega_n$ be alphabets. For any $\alpha_1 \in \Omega_1^\infty, \dots, \alpha_n \in \Omega_n^\infty$, we use $\alpha_1 \times \cdots \times \alpha_n$ to denote an infinite sequence α over $\Omega_1 \times \cdots \times \Omega_n$ such that $\alpha(k) = (\alpha_1(k), \dots, \alpha_n(k))$ for every $k \in \mathbb{N}^+$. For any $\sigma_1 \in \Omega_1^*, \dots, \sigma_n \in \Omega_n^*$ with $|\sigma_1| = \cdots = |\sigma_n|$, we define $\sigma_1 \times \cdots \times \sigma_n$ in a similar manner.

Definition 23 (Independence of ensembles). *Let $\Omega_1, \dots, \Omega_n$ be alphabets, and let $P_1 \in \mathbb{P}(\Omega_1), \dots, P_n \in \mathbb{P}(\Omega_n)$. Let $\alpha_1, \dots, \alpha_n$ be ensembles for P_1, \dots, P_n , respectively. We say that $\alpha_1, \dots, \alpha_n$ are independent if $\alpha_1 \times \cdots \times \alpha_n$ is an ensemble for $P_1 \times \cdots \times P_n$.* \square

Note that the notion of the independence of ensembles in our theory corresponds to the notion of *independence* of collectives in the theory of collectives by von Mises [31]. Theorem 25 below gives equivalent characterizations of the notion of the independence of random variables in terms of that of ensembles. To prove Theorem 25, we first show the following proposition.

Proposition 24. *Let $\alpha \in \Omega^\infty$, and let $X_1: \Omega \rightarrow \Omega_1, \dots, X_n: \Omega \rightarrow \Omega_n$ be random variables on Ω . Then $(X_1 \times \cdots \times X_n)(\alpha) = X_1(\alpha) \times \cdots \times X_n(\alpha)$.*

Proof. For each $k \in \mathbb{N}^+$, we see that

$$\begin{aligned}((X_1 \times \cdots \times X_n)(\alpha))(k) &= (X_1 \times \cdots \times X_n)(\alpha(k)) \\ &= (X_1(\alpha(k)), \dots, X_n(\alpha(k))) \\ &= ((X_1(\alpha))(k), \dots, (X_n(\alpha))(k)) \\ &= (X_1(\alpha) \times \cdots \times X_n(\alpha))(k).\end{aligned}$$

This completes the proof. \square

Theorem 25. *Let $P \in \mathbb{P}(\Omega)$, and let $X_1: \Omega \rightarrow \Omega_1, \dots, X_n: \Omega \rightarrow \Omega_n$ be random variables on Ω . Then the following conditions are equivalent to one another.*

- (i) *The random variables X_1, \dots, X_n are independent on P .*
- (ii) *For every ensemble α for P , the ensembles $X_1(\alpha), \dots, X_n(\alpha)$ are independent.*
- (iii) *There exists an ensemble α for P such that the ensembles $X_1(\alpha), \dots, X_n(\alpha)$ are independent.*

Proof. Assume that the condition (i) holds. Let α be an arbitrary ensemble for the finite probability space P . It follows from Theorem 22 that $X_i(\alpha)$ is Martin-Löf $X_i(P)$ -random for every $i = 1, 2, \dots, n$, and $(X_1 \times \dots \times X_n)(\alpha)$ is Martin-Löf $(X_1 \times \dots \times X_n)(P)$ -random. Therefore, by Proposition 24 and Proposition 21, we see that $X_1(\alpha) \times \dots \times X_n(\alpha)$ is Martin-Löf $X_1(P) \times \dots \times X_n(P)$ -random. Hence, we have the implication (i) \Rightarrow (ii).

Since there exists an ensemble α for the finite probability space P by Theorem 7, the implication (ii) \Rightarrow (iii) is obvious.

Finally, the implication (iii) \Rightarrow (i) is shown as follows. Assume that the condition (iii) holds. Then it follows from Proposition 24 that $(X_1 \times \dots \times X_n)(\alpha)$ is Martin-Löf $X_1(P) \times \dots \times X_n(P)$ -random for some ensemble α for the finite probability space P . On the other hand, by Theorem 22 we see that $(X_1 \times \dots \times X_n)(\alpha)$ is Martin-Löf $(X_1 \times \dots \times X_n)(P)$ -random. It follows from Corollary 14 we that $X_1(P) \times \dots \times X_n(P) = (X_1 \times \dots \times X_n)(P)$. Thus, by Proposition 21 we have that X_1, \dots, X_n are independent on P . This completes the proof. \square

Next, we consider the operational characterizations of the notion of the independence of an arbitrary number of events, in terms of ensembles. Let A_1, \dots, A_n be events on a finite probability space $P \in \mathbb{P}(\Omega)$. We say that the events A_1, \dots, A_n are *independent on P* if for every i_1, \dots, i_k with $1 \leq i_1 < \dots < i_k \leq n$ it holds that

$$P(A_{i_1} \cap \dots \cap A_{i_k}) = P(A_{i_1}) \dots P(A_{i_k}).$$

For any $A \subset \Omega$, we use χ_A to denote a function $f: \Omega \rightarrow \{0, 1\}$ such that $f(a) := 1$ if $a \in A$ and $f(a) := 0$ otherwise. Note that $C_A(\alpha) = \chi_A(\alpha)$ for every $A \subset \Omega$ and $\alpha \in \Omega^\infty$. It is then easy to show the following proposition.

Proposition 26. *Let $P \in \mathbb{P}(\Omega)$, and let $A_1, \dots, A_n \subset \Omega$. Then the events A_1, \dots, A_n are independent on P if and only if the random variables $\chi_{A_1}, \dots, \chi_{A_n}$ are independent on P .* \square

Using Proposition 26, Theorem 25 results in Theorem 27 below, which gives equivalent characterizations of the notion of the independence of an arbitrary number of events in terms of that of ensembles.

Theorem 27. *Let A_1, \dots, A_n be events on a finite probability space $P \in \mathbb{P}(\Omega)$. Then the following conditions are equivalent to one another.*

- (i) *The events A_1, \dots, A_n are independent on P .*
- (ii) *For every ensemble α for P , the ensembles $C_{A_1}(\alpha), \dots, C_{A_n}(\alpha)$ are independent.*
- (iii) *There exists an ensemble α for P such that the ensembles $C_{A_1}(\alpha), \dots, C_{A_n}(\alpha)$ are independent.* \square

8 Further equivalence of the notions of independence on computable finite probability spaces

In the preceding section we saw that the independence of an arbitrary number of events/random variables and that of ensembles are equivalent to each other on an arbitrary finite probability space. In this section we show that these independence notions are further equivalent to the notion of the independence in the sense of van Lambalgen's Theorem [28] in the case where the underlying finite probability space is *computable*. Thus, *the three independence notions are equivalent to one another* in this case. To show the equivalence, we generalize van Lambalgen's Theorem [28] over our framework first.

8.1 A generalization of van Lambalgen's Theorem

To study a generalization of van Lambalgen's Theorem, first we generalize the notion of Martin-Löf P -randomness over *relativized computation* and introduce the notion of *Martin-Löf P -randomness relative to an oracle*.

The *relativized computation* is a generalization of normal computation. Let $\beta_1, \dots, \beta_\ell$ be arbitrary infinite sequences over an alphabet. In the relativized computation, a (deterministic) Turing machine is allowed to refer to $\beta_1, \dots, \beta_\ell$ as an *oracle* during the computation. Namely, in the relativized computation, a Turing machine can query $(k, n) \in \{1, \dots, \ell\} \times \mathbb{N}^+$ at any time and then obtains the response $\beta_k(n)$ during the computation. Such a Turing machine is called an *oracle Turing machine*. The relativized computation is more powerful than normal computation, in general.

We define the notion of a *Martin-Löf P -test relative to $\beta_1, \dots, \beta_\ell$* as a Martin-Löf P -test where the Turing machine computing the Martin-Löf P -test is an oracle Turing machine which can refer to the sequences $\beta_1, \dots, \beta_\ell$ during the computation. Based on this notion, we define the notion of *Martin-Löf P -randomness relative to $\beta_1, \dots, \beta_\ell$* in the same manner as (ii) and (iii) of Definition 6. Formally, the notion of *Martin-Löf P -randomness relative to infinite sequences* is defined as follows.

Definition 28 (Martin-Löf P -randomness relative to infinite sequences). *Let $P \in \mathbb{P}(\Omega)$, and let $\beta_1, \dots, \beta_\ell$ be infinite sequences over an alphabet. A subset \mathcal{C} of $\mathbb{N}^+ \times \Omega^*$ is called a Martin-Löf P -test relative to $\beta_1, \dots, \beta_\ell$ if the following holds.*

(i) *There exists an oracle Turing machine \mathcal{M} such that*

$$\mathcal{C} = \{x \in \mathbb{N}^+ \times \Omega^* \mid \mathcal{M} \text{ accepts } x \text{ relative to } \beta_1, \dots, \beta_\ell\};$$

(ii) *For every $n \in \mathbb{N}^+$ it holds that $\lambda_P([\mathcal{C}_n]^\prec) < 2^{-n}$ where $\mathcal{C}_n := \{ \sigma \mid (n, \sigma) \in \mathcal{C} \}$.*

For any $\alpha \in \Omega^\infty$, we say that α is Martin-Löf P -random relative to $\beta_1, \dots, \beta_\ell$ if for every Martin-Löf P -test \mathcal{C} relative to $\beta_1, \dots, \beta_\ell$ there exists $n \in \mathbb{N}^+$ such that $\alpha \notin [\mathcal{C}_n]^\prec$. \square

Obviously, the following holds.

Proposition 29. *Let $P \in \mathbb{P}(\Omega)$, and let $\beta_1, \dots, \beta_\ell$ be infinite sequences over an alphabet. For every $\alpha \in \Omega^\infty$, if α is Martin-Löf P -random relative to $\beta_1, \dots, \beta_\ell$ then α is Martin-Löf P -random. \square*

The converse does not necessarily hold. In the case where α is Martin-Löf P -random, the converse means that the Martin-Löf P -randomness of α is *independent* of $\beta_1, \dots, \beta_\ell$ in a certain sense.

Let β be an infinite sequence over an alphabet. For any $\alpha \in \{0, 1\}^\infty$, we say that α is *Martin-Löf random relative to β* if α is Martin-Löf U -random relative to β where $U \in \mathbb{P}(\{0, 1\})$ such that $U(0) = U(1) = 1/2$. Based on this notion of *Martin-Löf randomness relative to an infinite sequence*, van Lambalgen's Theorem is stated as follows.

Theorem 30 (van Lambalgen's Theorem, van Lambalgen [28]). *Let $\alpha, \beta \in \{0, 1\}^\infty$, and let $\alpha \oplus \beta$ denote the infinite binary sequence*

$$\alpha(1)\beta(1)\alpha(2)\beta(2)\alpha(3)\beta(3)\dots\dots\dots$$

Then the following conditions are equivalent.

(i) $\alpha \oplus \beta$ is Martin-Löf random.

(ii) α is Martin-Löf random relative to β and β is Martin-Löf random. □

We generalize van Lambalgen's Theorem as follows.

Theorem 31 (Generalization of van Lambalgen's Theorem I). *Let Ω_1 and Ω_2 be alphabets, and let $P_1 \in \mathbb{P}(\Omega_1)$ and $P_2 \in \mathbb{P}(\Omega_2)$. Let $\alpha_1 \in \Omega_1^\infty$ and $\alpha_2 \in \Omega_2^\infty$, and let $\beta_1, \dots, \beta_\ell$ be infinite sequences over an alphabet. Suppose that P_1 is computable. Then $\alpha_1 \times \alpha_2$ is Martin-Löf $P_1 \times P_2$ -random relative to $\beta_1, \dots, \beta_\ell$ if and only if α_1 is Martin-Löf P_1 -random relative to $\alpha_2, \beta_1, \dots, \beta_\ell$ and α_2 is Martin-Löf P_2 -random relative to $\beta_1, \dots, \beta_\ell$. □*

The proof of Theorem 31 is obtained by generalizing and elaborating the proof of van Lambalgen's Theorem given in Nies [16, Section 3.4]. The detail is given in the subsequent two subsections. Note that in Theorem 31, the computability of P_1 is assumed while that of P_2 is not required.

We have Theorem 32 below based on Theorem 31. Note that the computability of P_n is not required in Theorem 32.

Theorem 32 (Generalization of van Lambalgen's Theorem II). *Let $n \geq 2$. Let $\Omega_1, \dots, \Omega_n$ be alphabets, and let $P_1 \in \mathbb{P}(\Omega_1), \dots, P_n \in \mathbb{P}(\Omega_n)$. For each $i = 1, \dots, n$, let $\alpha_i \in \Omega_i^\infty$, and let $\beta_1, \dots, \beta_\ell$ be infinite sequences over an alphabet. Suppose that P_1, \dots, P_{n-1} are computable. Then $\alpha_1 \times \dots \times \alpha_n$ is Martin-Löf $P_1 \times \dots \times P_n$ -random relative to $\beta_1, \dots, \beta_\ell$ if and only if for every $k = 1, \dots, n$ it holds that α_k is Martin-Löf P_k -random relative to $\alpha_{k+1}, \dots, \alpha_n, \beta_1, \dots, \beta_\ell$.*

Proof. We show the result by induction on $n \geq 2$. In the case of $n = 2$, the result holds since it is precisely Theorem 31.

For an arbitrary $m \geq 2$, assume that the result holds for $n = m$. Let $\Omega_1, \dots, \Omega_{m+1}$ be alphabets, and let $P_1 \in \mathbb{P}(\Omega_1), \dots, P_{m+1} \in \mathbb{P}(\Omega_{m+1})$. For each $i = 1, \dots, m+1$, let $\alpha_i \in \Omega_i^\infty$, and let $\beta_1, \dots, \beta_\ell$ be infinite sequences over an alphabet. Suppose that P_1, \dots, P_m are computable. Then, by applying Theorem 31 with $P_1 \times \dots \times P_m$ as P_1 , P_{m+1} as P_2 , $\alpha_1 \times \dots \times \alpha_m$ as α_1 , and α_{m+1} as α_2 , we have that $(\alpha_1 \times \dots \times \alpha_m) \times \alpha_{m+1}$ is Martin-Löf $(P_1 \times \dots \times P_m) \times P_{m+1}$ -random relative to $\beta_1, \dots, \beta_\ell$ if and only if $\alpha_1 \times \dots \times \alpha_m$ is Martin-Löf $P_1 \times \dots \times P_m$ -random relative to $\alpha_{m+1}, \beta_1, \dots, \beta_\ell$ and α_{m+1} is Martin-Löf P_{m+1} -random relative to $\beta_1, \dots, \beta_\ell$. Thus, by applying the result for $n = m$ we have the result for $n = m + 1$. This completes the proof. □

8.2 The proof of the “only if” part of Theorem 31

We prove the following theorem, from which the “only if” part of Theorem 31 follows.

Theorem 33. *Let Ω_1 and Ω_2 be alphabets, and let $P_1 \in \mathbb{P}(\Omega_1)$ and $P_2 \in \mathbb{P}(\Omega_2)$. Let $\alpha_1 \in \Omega_1^\infty$ and $\alpha_2 \in \Omega_2^\infty$, and let $\beta_1, \dots, \beta_\ell$ be infinite sequences over an alphabet Θ . Suppose that P_1 is right-computable. If $\alpha_1 \times \alpha_2$ is Martin-Löf $P_1 \times P_2$ -random relative to $\beta_1, \dots, \beta_\ell$ then α_1 is Martin-Löf P_1 -random relative to $\alpha_2, \beta_1, \dots, \beta_\ell$ and α_2 is Martin-Löf P_2 -random relative to $\beta_1, \dots, \beta_\ell$. \square*

In order to prove Theorem 33, we use the notion of *universal Martin-Löf P -test relative to infinite sequences*.

Definition 34 (Universal Martin-Löf P -test relative to infinite sequences). *Let $P \in \mathbb{P}(\Omega)$, and let Θ be an alphabet. Let $\ell \in \mathbb{N}^+$. An oracle Turing machine \mathcal{M} is called a universal Martin-Löf P -test relative to ℓ infinite sequences over Θ if for every $\beta_1, \dots, \beta_\ell \in \Theta^\infty$ there exists \mathcal{C} such that*

- (i) $\mathcal{C} = \{x \in \mathbb{N}^+ \times \Omega^* \mid \mathcal{M} \text{ accepts } x \text{ relative to } \beta_1, \dots, \beta_\ell\}$,
- (ii) for every $n \in \mathbb{N}^+$ it holds that $\lambda_P([C_n]^\prec) < 2^{-n}$ where $C_n := \{\sigma \mid (n, \sigma) \in \mathcal{C}\}$, and
- (iii) for every Martin-Löf P -test \mathcal{D} relative to $\beta_1, \dots, \beta_\ell$,

$$\bigcap_{n=1}^{\infty} [\mathcal{D}_n]^\prec \subset \bigcap_{n=1}^{\infty} [C_n]^\prec.$$

\square

It is then easy to show the following theorem.

Theorem 35. *Let $P \in \mathbb{P}(\Omega)$, and let Θ be an alphabet. Let $\ell \in \mathbb{N}^+$. Suppose that P is right-computable. Then there exists a universal Martin-Löf P -test relative to ℓ infinite sequences over Θ . \square*

In a similar manner to the proof of Theorem 13 we can show the following theorem in the context of relativized computation.

Theorem 36. *Let $P \in \mathbb{P}(\Omega)$, and let $\beta_1, \dots, \beta_\ell$ be infinite sequences over an alphabet. Let $\alpha \in \Omega^\infty$, and let a and b be distinct elements of Ω . Suppose that γ is an infinite sequence over $\Omega \setminus \{b\}$ obtained by replacing all occurrences of b by a in α . If α is Martin-Löf P -random relative to $\beta_1, \dots, \beta_\ell$ then γ is Martin-Löf Q -random relative to $\beta_1, \dots, \beta_\ell$, where $Q \in \mathbb{P}(\Omega \setminus \{b\})$ such that $Q(x) := P(a) + P(b)$ if $x = a$ and $Q(x) := P(x)$ otherwise. \square*

Theorem 33 is then proved as follows, using Theorems 36 and 35.

Proof of Theorem 33. First, we show that if $\alpha_1 \times \alpha_2$ is Martin-Löf $P_1 \times P_2$ -random relative to $\beta_1, \dots, \beta_\ell$ then α_2 is Martin-Löf P_1 -random relative to $\beta_1, \dots, \beta_\ell$. This is easily shown using Theorem 36 repeatedly.

Next, we show that if $\alpha_1 \times \alpha_2$ is Martin-Löf $P_1 \times P_2$ -random relative to $\beta_1, \dots, \beta_\ell$ then α_1 is Martin-Löf P_1 -random relative to $\alpha_2, \beta_1, \dots, \beta_\ell$. Since P_1 is right-computable, it follows from Theorem 35 that there exists a universal Martin-Löf P_1 -test relative to $\ell + 1$ infinite sequences over $\Omega_2 \cup \Theta$. Thus, there exists an oracle Turing machine \mathcal{M} such that for every $\gamma \in \Omega_2^\infty$ there exists \mathcal{C} such that

- (i) $\mathcal{C} = \{x \in \mathbb{N}^+ \times \Omega_1^* \mid \mathcal{M} \text{ accepts } x \text{ relative to } \gamma, \beta_1, \dots, \beta_\ell\},$
- (ii) for every $n \in \mathbb{N}^+, \lambda_{P_1}([\mathcal{C}_n]^\prec) < 2^{-n},$ and
- (iii) for every Martin-Löf P_1 -test \mathcal{D} relative to $\gamma, \beta_1, \dots, \beta_\ell,$

$$\bigcap_{n=1}^{\infty} [\mathcal{D}_n]^\prec \subset \bigcap_{n=1}^{\infty} [\mathcal{C}_n]^\prec.$$

For each $\sigma \in \Omega_2^*$, let \mathcal{U}^σ be the set of all $x \in \mathbb{N}^+ \times \Omega_1^*$ such that \mathcal{M} accepts x relative to $\sigma 0^\infty, \beta_1, \dots, \beta_\ell$ with oracle access only to the prefix of $\sigma 0^\infty$ of length $|\sigma|$ in the first infinite sequence. Here, $\sigma 0^\infty$ denotes the infinite sequence over $\Omega_2 \cup \{0\}$ which is the concatenation of the finite string σ and the infinite sequence consisting only of 0. It follows that $\lambda_{P_1}([\mathcal{U}_n^\sigma]^\prec) < 2^{-n}$ for every $n \in \mathbb{N}^+$ and every $\sigma \in \Omega_2^*$, where $\mathcal{U}_n^\sigma := \{\tau \mid (n, \tau) \in \mathcal{U}^\sigma\}$. For each $k, n \in \mathbb{N}^+$, let

$$G_n(k) = \{u \times \sigma \mid u \in \Omega_1^k \text{ \& \& } \sigma \in \Omega_2^k \text{ \& \& } \text{Some prefix of } u \text{ is in } \mathcal{U}_n^\sigma\}.$$

Then $G_n(k)$ is r.e. relative to $\beta_1, \dots, \beta_\ell$ uniformly in n and k . For each $n, k \in \mathbb{N}^+$, we see that

$$\lambda_{P_1 \times P_2}([G_n(k)]^\prec) \leq \sum_{\sigma \in \Omega_2^k} \lambda_{P_1}([\mathcal{U}_n^\sigma]^\prec) \lambda_{P_2}([\sigma]^\prec) < \sum_{\sigma \in \Omega_2^k} 2^{-n} \lambda_{P_2}([\sigma]^\prec) = 2^{-n}.$$

On the other hand, it follows that $[G_n(k)]^\prec \subset [G_n(k+1)]^\prec$ for every $n, k \in \mathbb{N}^+$. For each $n \in \mathbb{N}^+$, let $G_n = \bigcup_{k=1}^{\infty} G_n(k)$. Then G_n is r.e. relative to $\beta_1, \dots, \beta_\ell$ uniformly in n , and $\lambda_{P_1 \times P_2}([G_n]^\prec) \leq 2^{-n}$ for every n . Thus, the set $\{(n, \sigma) \mid n \in \mathbb{N}^+ \text{ \& \& } \sigma \in G_n\}$ is a Martin-Löf $P_1 \times P_2$ -test relative to $\beta_1, \dots, \beta_\ell$.

For arbitrary $\alpha_1 \in \Omega_1^\infty$ and $\alpha_2 \in \Omega_2^\infty$, assume that α_1 is not Martin-Löf P_1 -random relative to $\alpha_2, \beta_1, \dots, \beta_\ell$. Then there exists \mathcal{C} such that

- (i) $\mathcal{C} = \{x \in \mathbb{N}^+ \times \Omega_1^* \mid \mathcal{M} \text{ accepts } x \text{ relative to } \alpha_2, \beta_1, \dots, \beta_\ell\},$
- (ii) for every $n \in \mathbb{N}^+, \lambda_{P_1}([\mathcal{C}_n]^\prec) < 2^{-n},$ and
- (iii)

$$\alpha_1 \in \bigcap_{n=1}^{\infty} [\mathcal{C}_n]^\prec.$$

For each $n \in \mathbb{N}^+$, there exists $m \in \mathbb{N}^+$ such that $\alpha_1 \upharpoonright_m \in \mathcal{C}_n$. Then, there exists $k \geq m$ such that \mathcal{M} accepts $\alpha_1 \upharpoonright_m$ relative to $\alpha_2, \beta_1, \dots, \beta_\ell$ with oracle access only to the prefix of α_2 of length k in the first infinite sequence α_2 . It follows that $\alpha_1 \upharpoonright_m \in \mathcal{U}_n^{\alpha_2 \upharpoonright_k}$. Thus, $\alpha_1 \upharpoonright_k \times \alpha_2 \upharpoonright_k \in G_n(k)$, and therefore $\alpha_1 \times \alpha_2 \in [G_n(k)]^\prec \subset [G_n]^\prec$. Hence, $\alpha_1 \times \alpha_2$ is not Martin-Löf $P_1 \times P_2$ -random relative to $\beta_1, \dots, \beta_\ell$. This completes the proof. \square

8.3 The proof of the “if” part of Theorem 31

Next, we prove the following theorem, from which the “if” part of Theorem 31 follows.

Theorem 37. *Let Ω_1 and Ω_2 be alphabets, and let $P_1 \in \mathbb{P}(\Omega_1)$ and $P_2 \in \mathbb{P}(\Omega_2)$. Let $\alpha_1 \in \Omega_1^\infty$ and $\alpha_2 \in \Omega_2^\infty$, and let $\beta_1, \dots, \beta_\ell$ be infinite sequences over an alphabet. Suppose that P_1 is left-computable. If α_1 is Martin-Löf P_1 -random relative to $\alpha_2, \beta_1, \dots, \beta_\ell$ and α_2 is Martin-Löf P_2 -random relative to $\beta_1, \dots, \beta_\ell$, then $\alpha_1 \times \alpha_2$ is Martin-Löf $P_1 \times P_2$ -random relative to $\beta_1, \dots, \beta_\ell$.*

Proof. Suppose that $\alpha_1 \times \alpha_2$ is not Martin-Löf $P_1 \times P_2$ -random relative to $\beta_1, \dots, \beta_\ell$. Then there exists a Martin-Löf P -test \mathcal{V} relative to $\beta_1, \dots, \beta_\ell$ such that

- (i) \mathcal{V}_d is prefix-free for every $d \in \mathbb{N}^+$,
- (ii) $\lambda_{P_1 \times P_2}([\mathcal{V}_d]^\prec) < 2^{-2d}$ for every $d \in \mathbb{N}^+$, and
- (iii) $\alpha_1 \times \alpha_2 \in [\mathcal{V}_d]^\prec$ for every $d \in \mathbb{N}^+$.

On the one hand, for each $x \in \Omega_2^*$, we use $[\emptyset \times x]$ to denote the set

$$\{\gamma_1 \times \gamma_2 \mid \gamma_1 \in \Omega_1^\infty, \gamma_2 \in \Omega_2^\infty, \text{ and } x \text{ is a prefix of } \gamma_2\}.$$

On the other hand, for each $x \in \Omega_2^*$ and $W \subset (\Omega_1 \times \Omega_2)^*$, we use $F(W, x)$ to denote the set of all $\sigma_1 \in \Omega_1^*$ such that there exists $\sigma_2 \in \Omega_2^*$ for which (i) $|\sigma_1| = |\sigma_2|$, (ii) $\sigma_1 \times \sigma_2 \in W$, and (iii) σ_2 is a prefix of x . It is then easy to see that

$$P_1(F(W, x))P_2(x) = \lambda_{P_1 \times P_2}([W]^\prec \cap [\emptyset \times x]) \quad (8)$$

for every $x \in \Omega_2^*$ and every prefix-free subset W of $(\Omega_1 \times \Omega_2)^{\leq |x|}$. For each $d \in \mathbb{N}^+$, let

$$S_d = \{x \in \Omega_2^* \mid 2^{-d} < P_1(F(\mathcal{V}_d \cap (\Omega_1 \times \Omega_2)^{\leq |x|}, x))\}.$$

Since P_1 is left-computable, S_d is r.e. relative to $\beta_1, \dots, \beta_\ell$ uniformly in d .

Let $d \in \mathbb{N}^+$. Let $\{x_i\}$ be a listing of the minimal strings in S_d . It follows that

$$\begin{aligned} 2^{-d} \lambda_{P_2}([x_i]^\prec) &= 2^{-d} P_2(x_i) \leq P_1(F(\mathcal{V}_d \cap (\Omega_1 \times \Omega_2)^{\leq |x_i|}, x_i)) P_2(x_i) \\ &= \lambda_{P_1 \times P_2} \left(\left[\mathcal{V}_d \cap (\Omega_1 \times \Omega_2)^{\leq |x_i|} \right]^\prec \cap [\emptyset \times x_i] \right) \\ &\leq \lambda_{P_1 \times P_2}([\mathcal{V}_d]^\prec \cap [\emptyset \times x_i]). \end{aligned}$$

Since the sets $[\mathcal{V}_d]^\prec \cap [\emptyset \times x_i]$ are pairwise disjoint, we have

$$\sum_i 2^{-d} \lambda_{P_2}([x_i]^\prec) \leq \sum_i \lambda_{P_1 \times P_2}([\mathcal{V}_d]^\prec \cap [\emptyset \times x_i]) \leq \lambda_{P_1 \times P_2}([\mathcal{V}_d]^\prec) < 2^{-2d}$$

for each $d \in \mathbb{N}^+$. Therefore $\lambda_{P_2}([S_d]^\prec) = \sum_i \lambda_{P_2}([x_i]^\prec) < 2^{-d}$ for each $d \in \mathbb{N}^+$. It follows that

$$\lambda_{P_2} \left(\left[\bigcup_{c=d}^{\infty} S_{c+1} \right]^\prec \right) \leq \sum_{c=d}^{\infty} \lambda_{P_2}([S_{c+1}]^\prec) < 2^{-d}$$

for each $d \in \mathbb{N}^+$, and $\bigcup_{c=d}^{\infty} S_{c+1}$ is r.e. relative to $\beta_1, \dots, \beta_\ell$ uniformly in d . Hence, the set $\{(n, \sigma) \mid n \in \mathbb{N}^+ \text{ \& } \sigma \in \bigcup_{c=d}^{\infty} S_{c+1}\}$ is a Martin-Löf P_2 -test relative to $\beta_1, \dots, \beta_\ell$.

In the case where $\alpha_2 \in [S_d]^\prec$ for infinitely many d , we have that $\alpha_2 \in [\bigcup_{c=d}^{\infty} S_{c+1}]^\prec$ for every d , and therefore α_2 is not Martin-Löf P_2 -random relative to $\beta_1, \dots, \beta_\ell$. Thus, the theorem holds in this case. Therefore, in what follows we assume that there exists $d_0 \in \mathbb{N}^+$ such that $\alpha_2 \notin [S_d]^\prec$ for every $d > d_0$. We will then show that α_1 is not Martin-Löf P_1 -random relative to $\alpha_2, \beta_1, \dots, \beta_\ell$.

For each $d, n \in \mathbb{N}^+$, let

$$H_d(n) = \{w \in \Omega_1^n \mid [w \times \alpha_2 \upharpoonright_n]^\prec \subset [\mathcal{V}_d \cap (\Omega_1 \times \Omega_2)^{\leq n}]^\prec\}.$$

Let $d, n \in \mathbb{N}^+$, and let w_1, \dots, w_m be a listing of all elements of $H_d(n)$. Since

$$[w_i \times \alpha_2 \upharpoonright_n]^\prec \subset [\mathcal{V}_d \cap (\Omega_1 \times \Omega_2)^{\leq n}]^\prec \cap [\emptyset \times \alpha_2 \upharpoonright_n]$$

for every $i = 1, \dots, m$, and the sets $[w_i \times \alpha_2 \upharpoonright_n]^\prec$ are pairwise disjoint, we see that

$$\begin{aligned} \lambda_{P_1}([H_d(n)]^\prec) \lambda_{P_2}([\alpha_2 \upharpoonright_n]^\prec) &= \left(\sum_{i=1}^m \lambda_{P_1}([w_i]^\prec) \right) \lambda_{P_2}([\alpha_2 \upharpoonright_n]^\prec) \\ &= \sum_{i=1}^m \lambda_{P_1}([w_i]^\prec) \lambda_{P_2}([\alpha_2 \upharpoonright_n]^\prec) \\ &= \sum_{i=1}^m \lambda_{P_1 \times P_2}([w_i \times \alpha_2 \upharpoonright_n]^\prec) \\ &\leq \lambda_{P_1 \times P_2}([\mathcal{V}_d \cap (\Omega_1 \times \Omega_2)^{\leq n}]^\prec \cap [\emptyset \times \alpha_2 \upharpoonright_n]). \end{aligned} \tag{9}$$

Assume that $d > d_0$. Then, since $\alpha_2 \notin [S_d]^\prec$ we have $\alpha_2 \upharpoonright_n \notin S_d$. It follows from (8) that

$$\begin{aligned} \lambda_{P_1 \times P_2}([\mathcal{V}_d \cap (\Omega_1 \times \Omega_2)^{\leq n}]^\prec \cap [\emptyset \times \alpha_2 \upharpoonright_n]) &= P_1(F(\mathcal{V}_d \cap (\Omega_1 \times \Omega_2)^{\leq n}, \alpha_2 \upharpoonright_n)) P_2(\alpha_2 \upharpoonright_n) \\ &\leq 2^{-d} \lambda_{P_2}([\alpha_2 \upharpoonright_n]^\prec). \end{aligned}$$

Therefore, using (9) we have

$$\lambda_{P_1}([H_d(n)]^\prec) \lambda_{P_2}([\alpha_2 \upharpoonright_n]^\prec) \leq 2^{-d} \lambda_{P_2}([\alpha_2 \upharpoonright_n]^\prec).$$

Since α_2 is Martin-Löf P_2 -random relative to $\beta_1, \dots, \beta_\ell$, we can show that $\lambda_{P_2}([\alpha_2 \upharpoonright_n]^\prec) > 0$, in a similar manner to the proof of Theorem 10. Hence, we see that

$$\lambda_{P_1}([H_d(n)]^\prec) \leq 2^{-d} \tag{10}$$

for every $d > d_0$ and n .

On the other hand, we see that $[H_d(n)]^\prec \subset [H_d(n+1)]^\prec$ for every d and n . For each $d \geq d_0$, let $H_d = \bigcup_{n=1}^{\infty} H_d(n)$. It follows from (10) that $\lambda_{P_1}([H_d]^\prec) \leq 2^{-d}$ for every $d > d_0$. It is also easy to see that H_d is r.e. relative to $\alpha_2, \beta_1, \dots, \beta_\ell$ uniformly in d . Hence, the set $\{(n, \sigma) \mid n \in \mathbb{N}^+ \text{ \& } \sigma \in H_{n+d_0}\}$ is a Martin-Löf P_1 -test relative to $\alpha_2, \beta_1, \dots, \beta_\ell$.

Let $d \in \mathbb{N}^+$. Since $\alpha_1 \times \alpha_2 \in [\mathcal{V}_d]^\prec$, there exists $n \in \mathbb{N}^+$ such that $\alpha_1 \times \alpha_2 \upharpoonright_n \in \mathcal{V}_d$. It follows that $\alpha_1 \upharpoonright_n \times \alpha_2 \upharpoonright_n \in \mathcal{V}_d \cap (\Omega_1 \times \Omega_2)^{\leq n}$, and therefore $\alpha_1 \upharpoonright_n \in H_d(n)$. It follows that $\alpha_1 \in [H_d(n)]^\prec \subset [H_d]^\prec$. Therefore, $\alpha_1 \in [H_d]^\prec$ for every $d \in \mathbb{N}^+$. Hence, α_1 is not Martin-Löf P_1 -random relative to $\alpha_2, \beta_1, \dots, \beta_\ell$. This completes the proof. \square

8.4 Equivalence between the three independence notions on computable finite probability spaces

Theorem 38 below gives an equivalent characterization of the notion of the independence of ensembles in terms of Martin-Löf P -randomness relative to an oracle.

Theorem 38 (Generalization of van Lambalgen's Theorem III). *Let $n \geq 2$. Let $\Omega_1, \dots, \Omega_n$ be alphabets, and let $P_1 \in \mathbb{P}(\Omega_1), \dots, P_n \in \mathbb{P}(\Omega_n)$. Let $\alpha_1, \dots, \alpha_n$ be ensembles for P_1, \dots, P_n , respectively. Suppose that P_1, \dots, P_{n-1} are computable.⁴ Then the ensembles $\alpha_1, \dots, \alpha_n$ are independent if and only if for every $k = 1, \dots, n-1$ it holds that α_k is Martin-Löf P_k -random relative to $\alpha_{k+1}, \dots, \alpha_n$. \square*

Proof. Theorem 38 follows immediately from Theorem 32. \square

Combining Theorem 25 with Theorem 38 we obtain the following theorem.

Theorem 39. *Let $P \in \mathbb{P}(\Omega)$, and let $X_1: \Omega \rightarrow \Omega_1, \dots, X_n: \Omega \rightarrow \Omega_n$ be random variables on Ω . Suppose that $X_1(P), \dots, X_{n-1}(P)$ are computable. Then the following conditions are equivalent to one another.*

- (i) *The random variables X_1, \dots, X_n are independent on P .*
- (ii) *For every ensemble α for P and every $k = 1, \dots, n-1$ it holds that $X_k(\alpha)$ is Martin-Löf $X_k(P)$ -random relative to $X_{k+1}(\alpha), \dots, X_n(\alpha)$.*
- (iii) *There exists an ensemble α for P such that for every $k = 1, \dots, n-1$ it holds that $X_k(\alpha)$ is Martin-Löf $X_k(P)$ -random relative to $X_{k+1}(\alpha), \dots, X_n(\alpha)$.*

Proof. Let α be an arbitrary ensemble for P . Then it follows from Theorem 22 that $X_1(\alpha), \dots, X_n(\alpha)$ are ensembles for $X_1(P), \dots, X_n(P)$, respectively. Therefore, in the case where $X_1(P), \dots, X_{n-1}(P)$ are computable, using Theorem 38 we have that the ensembles $X_1(\alpha), \dots, X_n(\alpha)$ are independent if and only if for every $k = 1, \dots, n-1$ it holds that $X_k(\alpha)$ is Martin-Löf $X_k(P)$ -random relative to $X_{k+1}(\alpha), \dots, X_n(\alpha)$. Thus, Theorem 39 follows from Theorem 25. \square

Theorem 25 and Theorem 39 together show that the three independence notions we have considered so far: the independence of random variables, the independence of ensembles, and the independence in the sense of van Lambalgen's Theorem, are equivalent to one another on an arbitrary computable finite probability space.

Theorem 39 results in Theorem 40 below. Theorem 27 and Theorem 40 together show that the three independence notions are equivalent for arbitrary events, instead of random variables, on an arbitrary computable finite probability space.

Theorem 40. *Let A_1, \dots, A_n be events on a finite probability space $P \in \mathbb{P}(\Omega)$. Suppose that the finite probability space $C(P, A_k)$ is computable for every $k = 1, \dots, n-1$. Then the following conditions are equivalent to one another.*

- (i) *The events A_1, \dots, A_n are independent on P .*

⁴The computability of P_n is not required in the theorem.

(ii) For every ensemble α for P and every $k = 1, \dots, n-1$ it holds that $C_{A_k}(\alpha)$ is Martin-Löf $C(P, A_k)$ -random relative to $C_{A_{k+1}}(\alpha), \dots, C_{A_n}(\alpha)$.

(iii) There exists an ensemble α for P such that for every $k = 1, \dots, n-1$ it holds that $C_{A_k}(\alpha)$ is Martin-Löf $C(P, A_k)$ -random relative to $C_{A_{k+1}}(\alpha), \dots, C_{A_n}(\alpha)$.

Proof. The result is obtained by applying Theorem 39 to the random variables $\chi_{A_1}, \dots, \chi_{A_n}$ as X_1, \dots, X_n , respectively, and then using Proposition 26. \square

9 Applications

In this section we make applications of our framework to the general areas of science and technology in order to demonstrate the wide applicability of our framework to them. We here adopt information theory and cryptography as examples of the fields for the applications. Furthermore, we mention an application of our framework to quantum mechanics, which is developed in a series of works [21, 22, 23, 24, 25, 26, 27].

9.1 Application to information theory

In this subsection, we make an application of our framework to information theory [19]. *Instantaneous codes* play a basic role in the *noiseless source coding problem* in information theory, as described in what follows. See e.g., Ash [1] for the detail of the noiseless source coding by instantaneous codes.

Let Ω be an alphabet, as in the preceding sections. An instantaneous code C for Ω is an injective mapping from Ω to $\{0, 1\}^*$ such that $C(\Omega) := \{C(a) \mid a \in \Omega\}$ is a prefix-free set. A finite sequence $a_1, a_2, \dots, a_N \in \Omega$ is called a *message*. On the other hand, the finite binary string $C(a_1)C(a_2) \dots C(a_N)$ is called the *coded message* for a message a_1, a_2, \dots, a_N .

Let $P \in \mathbb{P}(\Omega)$ be a finite probability space. In the terminology of information theory, consider “independent identically distributed random variables X_1, X_2, \dots, X_N drawn from the probability mass function $P(a)$ with $a \in \Omega$.” In our framework, this means that we consider an alphabet Θ_N and a finite probability space $Q_N \in \mathbb{P}(\Theta_N)$, and consider random variables $X_1, \dots, X_N: \Theta_N \rightarrow \Omega$ on Θ_N such that X_1, \dots, X_N are independent on Q_N and $(X_i(Q_N))(a) = P(a)$ for every $i = 1, \dots, N$ and every $a \in \Omega$. In the source coding problem, the finite probability space P is called an *information source* which emits a symbol in Ω . The objective of the noiseless source coding problem is “to minimize the length of the coded message for a message a_1, a_2, \dots, a_N generated by the random variables X_1, X_2, \dots, X_N as $N \rightarrow \infty$.” For that purpose, it is sufficient to consider the *average codeword length* $L_P(C)$ of an instantaneous code C for a finite probability space P defined by

$$L_P(C) := \sum_{a \in \Omega} P(a) |C(a)|$$

independently on the value of N . We can then show that $L_P(C) \geq H(P)$ for every instantaneous code C for Ω and every finite probability space $P \in \mathbb{P}(\Omega)$, where $H(P)$ is the *Shannon entropy* of P defined by

$$H(P) := - \sum_{a \in \Omega} P(a) \log_2 P(a).$$

Hence, *the Shannon entropy gives the data compression limit for the noiseless source coding problem based on instantaneous codes*. For this reason, it is important to consider the notion of absolute optimality of an instantaneous code, where we say that an instantaneous code C for Ω is *absolutely optimal* for a finite probability space $P \in \mathbb{P}(\Omega)$ if $L_P(C) = H(P)$.

As an application of our framework to the noiseless source coding problem in information theory, we regard a “typical” infinite sequence in Ω^∞ which is a realization of the infinite sequence of the random variables X_1, X_2, X_3, \dots as an ensemble for the finite probability space P . For any $\alpha \in \Omega^\infty$ we use $\text{Coded}_C(\alpha)$ to denote an infinite binary sequence

$$C(\alpha(1))C(\alpha(2))C(\alpha(3))\dots\dots\dots$$

We can then show the following theorem.

Theorem 41. *Let $P \in \mathbb{P}(\Omega)$, and let C be an instantaneous code for Ω . Suppose that α is an ensemble for P . Then the following conditions are equivalent to each other.*

- (i) *The instantaneous code C is absolutely optimal for the finite probability space P .*
- (ii) *$\text{Coded}_C(\alpha)$ is Martin-Löf random.*

Proof. We note that the instantaneous code C for Ω is absolutely optimal for the finite probability space P if and only if $P(a) = 2^{-|C(a)|}$ for every $a \in \Omega$. For any $\sigma \in \Omega^+$, we use $C(\sigma)$ to denote $C(\sigma(1))C(\sigma(2))\dots C(\sigma(|\sigma|))$.

First, we show the implication (i) \Rightarrow (ii). For that purpose, suppose that C is absolutely optimal for P . Then

$$P(a) = 2^{-|C(a)|} \tag{11}$$

for every $a \in \Omega$. Assume contrarily that $\text{Coded}_C(\alpha)$ is not Martin-Löf random. Then there exists a Martin-Löf test $\mathcal{S} \subset \mathbb{N}^+ \times \{0, 1\}^*$ such that $\text{Coded}_C(\alpha) \in [\mathcal{S}_n]^\prec$ for every $n \in \mathbb{N}^+$. For each $\sigma \in \{0, 1\}^+$, let $f(\sigma)$ be the set of all $\tau \in \Omega^+$ such that σ is a prefix of $C(\tau(1))\dots C(\tau(|\tau|))$. We then define \mathcal{T} to be a subset of $\mathbb{N}^+ \times \Omega^*$ such that $\mathcal{T}_n = \bigcup_{\sigma \in \mathcal{S}_n} f(\sigma)$ for every $n \in \mathbb{N}^+$. Note from (11) that $\sum_{a \in \Omega} 2^{-|C(a)|} = 1$. Let $\sigma \in \{0, 1\}^+$. Then the set of the minimal strings in $f(\sigma)$ is finite. Therefore we denote it by $\{x_1, \dots, x_k\}$. It follows that $[\sigma]^\prec = \bigcup_i [C(x_i)]^\prec$. We thus have

$$\lambda_P([f(\sigma)]^\prec) = \sum_i \lambda_P([x_i]^\prec) = \sum_i P(x_i) = \sum_i 2^{-|C(x_i)|} = \sum_i \mathcal{L}([C(x_i)]^\prec) = \mathcal{L}([\sigma]^\prec),$$

where the third equality follows from (11), and \mathcal{L} is Lebesgue measure on $\{0, 1\}^\infty$. Therefore, we have that $\lambda_P([\mathcal{T}_n]^\prec) = \mathcal{L}([\mathcal{S}_n]^\prec) < 2^{-n}$ for each $n \in \mathbb{N}^+$. Since \mathcal{S} is r.e., \mathcal{T} is also r.e. Thus, \mathcal{T} is Martin-Löf P -test. On the other hand, it follows that $\alpha \in [\mathcal{T}_n]^\prec$ for every $n \in \mathbb{N}^+$. Hence, α is not Martin-Löf P -random. Thus we have a contradiction, and $\text{Coded}_C(\alpha)$ is Martin-Löf random.

Next, we show the implication (ii) \Rightarrow (i). We choose any specific $b_0 \notin \Omega$ and define Φ to be $\Omega \cup \{b_0\}$. Since $C(\Omega)$ is prefix-free, the Kraft inequality $\sum_{a \in \Omega} 2^{-|C(a)|} \leq 1$ holds. Hence, we can define a finite probability space $Q \in \mathbb{P}(\Phi)$ with the property that $Q(x) := 2^{-|C(x)|}$ if $x \in \Omega$ and $Q(x) := 1 - \sum_{a \in \Omega} 2^{-|C(a)|}$ otherwise. We show that if α is not Martin-Löf Q -random then $\text{Coded}_C(\alpha)$ is not Martin-Löf random. Thus, assume that α is not Martin-Löf Q -random. Then there exists a Martin-Löf Q -test $\mathcal{S} \subset \mathbb{N}^+ \times \Phi$ such that $\alpha \in [\mathcal{S}_n]^\prec$ for every $n \in \mathbb{N}^+$. We define \mathcal{T} to be a subset of $\mathbb{N}^+ \times \Omega^*$ such that $\mathcal{T}_n = \{C(\sigma) \mid \sigma \in \mathcal{S}_n \cap \Omega^+\}$ for every $n \in \mathbb{N}^+$. Thus,

since $Q(a) = 2^{-|C(a)|}$ for every $a \in \Omega$, we have $\lambda_Q([\sigma]^\prec) = Q(\sigma) = 2^{-|C(\sigma)|} = \mathcal{L}([C(\sigma)]^\prec)$ for each $\sigma \in \Omega^+$. Therefore, it is easy to see that $\mathcal{L}([\mathcal{T}_n]^\prec) = \lambda_Q([\mathcal{S}_n \cap \Omega^+]^\prec) \leq \lambda_Q([\mathcal{S}_n]^\prec) < 2^{-n}$ for each $n \in \mathbb{N}^+$. Since \mathcal{S} is r.e., \mathcal{T} is also r.e. Thus, \mathcal{T} is Martin-Löf test. On the other hand, it follows that $\text{Coded}_C(\alpha) \in [\mathcal{T}_n]^\prec$ for every $n \in \mathbb{N}^+$. Hence, $\text{Coded}_C(\alpha)$ is not Martin-Löf random, as desired.

Recall that α is a Martin-Löf P -random infinite sequence over Ω by the assumption of the theorem. We define a finite probability space $P' \in \mathbb{P}(\Phi)$ by the condition that $P'(x) := P(x)$ if $x \in \Omega$ and $P'(x) := 0$ otherwise. It follows that α is a Martin-Löf P' -random infinite sequence over Φ . Suppose that the condition (ii) holds. Then α is a Martin-Löf Q -random infinite sequence over Φ , as we showed above. It follows from Corollary 14 that $P' = Q$, and therefore $P(a) = 2^{-|C(a)|}$ for every $a \in \Omega$. Hence, the condition (i) holds. This completes the proof. \square

Recall from Theorem 2 that Martin-Löf random sequences are precisely the infinite binary sequences which cannot be compressible any more. Thus, Theorem 41 rephrases in a sharp manner the basic result of the noiseless source coding problem that the Shannon entropy gives the data compression limit, in the form of our framework.

9.2 Application to cryptography

In this subsection, we make an application of our framework to cryptography. We present new equivalent characterizations of the notion of *perfect secrecy* in terms of our framework.

The notion of perfect secrecy was introduced by Shannon [20], and plays a basic role in modern cryptography. First, we review the definition of encryption schemes to which the notion of perfect secrecy is applied.

Definition 42 (Encryption scheme). *Let \mathcal{M} , \mathcal{K} , and \mathcal{C} be alphabets. An encryption scheme over a message space \mathcal{M} , a key space \mathcal{K} , and a ciphertext space \mathcal{C} is a tuple $\Pi = (P_{\text{key}}, \text{Enc}, \text{Dec})$ such that*

- (i) $P_{\text{key}} \in \mathbb{P}(\mathcal{K})$,
- (ii) $\text{Enc}: \mathcal{M} \times \mathcal{K} \rightarrow \mathcal{C}$,
- (iii) $\text{Dec}: \mathcal{C} \times \mathcal{K} \rightarrow \mathcal{M}$, and
- (iv) $\text{Dec}(\text{Enc}(m, k), k) = m$ for every $m \in \mathcal{M}$ and $k \in \mathcal{K}$. \square

Let $\Pi = (P_{\text{key}}, \text{Enc}, \text{Dec})$ be as in Definition 42, and let $P_{\text{msg}} \in \mathbb{P}(\mathcal{M})$. The finite probability space P_{msg} serves as a “probability distribution” over message space \mathcal{M} for the encryption scheme Π . We then define random variables $M_\Pi: \mathcal{M} \times \mathcal{K} \rightarrow \mathcal{M}$ and $C_\Pi: \mathcal{M} \times \mathcal{K} \rightarrow \mathcal{C}$ on $\mathcal{M} \times \mathcal{K}$ by $M_\Pi(m, k) := m$ and $C_\Pi(m, k) := \text{Enc}(m, k)$, respectively. The notion of perfect secrecy is then defined as follows.

Definition 43 (Perfect secrecy, Shannon [20]). *Let \mathcal{M} , \mathcal{K} , and \mathcal{C} be alphabets. Let $\Pi = (P_{\text{key}}, \text{Enc}, \text{Dec})$ be an encryption scheme over a message space \mathcal{M} , a key space \mathcal{K} , and a ciphertext space \mathcal{C} . The encryption scheme Π is perfectly secret if for every $P_{\text{msg}} \in \mathbb{P}(\mathcal{M})$ it holds that the random variables M_Π and C_Π are independent on $P_{\text{msg}} \times P_{\text{key}}$. \square*

We use $U_{\mathcal{M}}$ to denote “the uniform distribution over a message space \mathcal{M} ,” i.e., to denote a finite probability space in $\mathbb{P}(\mathcal{M})$ such that

$$U_{\mathcal{M}}(m) = \frac{1}{\#\mathcal{M}}$$

for every $m \in \mathcal{M}$. Note that $U_{\mathcal{M}}$ is a *computable* finite probability space since every rational is computable. Based on Theorems 25 and 39 we can show Theorems 44 and 45 below, which characterize the notion of perfect secrecy equivalently in terms of the notions of the independence of ensembles and Martin-Löf P -randomness relative to an oracle.

Theorem 44 (New equivalent characterizations of perfect secrecy I). *Let \mathcal{M} , \mathcal{K} , and \mathcal{C} be alphabets. Let $\Pi = (P_{\text{key}}, \text{Enc}, \text{Dec})$ be an encryption scheme over a message space \mathcal{M} , a key space \mathcal{K} , and a ciphertext space \mathcal{C} . Then the following conditions are equivalent to one another.*

- (i) *The encryption scheme Π is perfectly secret.*
- (ii) *For every $P_{\text{msg}} \in \mathbb{P}(\mathcal{M})$ and every ensemble α for $P_{\text{msg}} \times P_{\text{key}}$, the ensembles $M_{\Pi}(\alpha)$ and $C_{\Pi}(\alpha)$ are independent.*
- (iii) *For every $P_{\text{msg}} \in \mathbb{P}(\mathcal{M})$ there exists an ensemble α for $P_{\text{msg}} \times P_{\text{key}}$ such that the ensembles $M_{\Pi}(\alpha)$ and $C_{\Pi}(\alpha)$ are independent.*
- (iv) *For every computable $P_{\text{msg}} \in \mathbb{P}(\mathcal{M})$ and every ensemble α for $P_{\text{msg}} \times P_{\text{key}}$ it holds that $M_{\Pi}(\alpha)$ is Martin-Löf P_{msg} -random relative to $C_{\Pi}(\alpha)$.*
- (v) *For every computable $P_{\text{msg}} \in \mathbb{P}(\mathcal{M})$ there exists an ensemble α for $P_{\text{msg}} \times P_{\text{key}}$ such that $M_{\Pi}(\alpha)$ is Martin-Löf P_{msg} -random relative to $C_{\Pi}(\alpha)$.*
- (vi) *For every ensemble α for $U_{\mathcal{M}} \times P_{\text{key}}$ it holds that $M_{\Pi}(\alpha)$ is Martin-Löf $U_{\mathcal{M}}$ -random relative to $C_{\Pi}(\alpha)$.*
- (vii) *There exists an ensemble α for $U_{\mathcal{M}} \times P_{\text{key}}$ such that $M_{\Pi}(\alpha)$ is Martin-Löf $U_{\mathcal{M}}$ -random relative to $C_{\Pi}(\alpha)$. \square*

Theorem 45 (New equivalent characterizations of perfect secrecy II). *Let \mathcal{M} , \mathcal{K} , and \mathcal{C} be alphabets. Let $\Pi = (P_{\text{key}}, \text{Enc}, \text{Dec})$ be an encryption scheme over a message space \mathcal{M} , a key space \mathcal{K} , and a ciphertext space \mathcal{C} . Suppose that P_{key} is computable. Then the following conditions are equivalent to one another.*

- (i) *The encryption scheme Π is perfectly secret.*
- (ii) *For every computable $P_{\text{msg}} \in \mathbb{P}(\mathcal{M})$ and every ensemble α for $P_{\text{msg}} \times P_{\text{key}}$ it holds that $C_{\Pi}(\alpha)$ is Martin-Löf $C_{\Pi}(P_{\text{msg}} \times P_{\text{key}})$ -random relative to $M_{\Pi}(\alpha)$.*
- (iii) *For every computable $P_{\text{msg}} \in \mathbb{P}(\mathcal{M})$ there exists an ensemble α for $P_{\text{msg}} \times P_{\text{key}}$ such that $C_{\Pi}(\alpha)$ is Martin-Löf $C_{\Pi}(P_{\text{msg}} \times P_{\text{key}})$ -random relative to $M_{\Pi}(\alpha)$.*
- (iv) *For every ensemble α for $U_{\mathcal{M}} \times P_{\text{key}}$ it holds that $C_{\Pi}(\alpha)$ is Martin-Löf $C_{\Pi}(U_{\mathcal{M}} \times P_{\text{key}})$ -random relative to $M_{\Pi}(\alpha)$.*

(v) There exists an ensemble α for $U_{\mathcal{M}} \times P_{\text{key}}$ such that $C_{\Pi}(\alpha)$ is Martin-Löf $C_{\Pi}(U_{\mathcal{M}} \times P_{\text{key}})$ -random relative to $M_{\Pi}(\alpha)$. \square

In Theorem 45 the computability of P_{key} is assumed while it is not assumed in Theorem 44. Note, however, that the finite probability space P_{key} , which serves as a “probability distribution” over key space \mathcal{K} , is normally *computable* in modern cryptography.

In order to prove Theorems 44 and 45 we need the following lemma.

Lemma 46. *Let \mathcal{M} , \mathcal{K} , and \mathcal{C} be alphabets. Let $\Pi = (P_{\text{key}}, \text{Enc}, \text{Dec})$ be an encryption scheme over a message space \mathcal{M} , a key space \mathcal{K} , and a ciphertext space \mathcal{C} . Then the following conditions are equivalent to one another.*

- (i) *The encryption scheme Π is perfectly secret.*
- (ii) *For every computable $P_{\text{msg}} \in \mathbb{P}(\mathcal{M})$ it holds that the random variables M_{Π} and C_{Π} are independent on $P_{\text{msg}} \times P_{\text{key}}$.*
- (iii) *The random variables M_{Π} and C_{Π} are independent on $U_{\mathcal{M}} \times P_{\text{key}}$.*

Proof. The implication (i) \Rightarrow (ii) is obvious. Since $U_{\mathcal{M}}$ is computable, the implication (ii) \Rightarrow (iii) is also obvious. Thus, we show the implication (iii) \Rightarrow (i) in what follows. For that purpose, we first note that the following hold for every $P_{\text{msg}} \in \mathbb{P}(\mathcal{M})$, $m \in \mathcal{M}$, and $c \in \mathcal{C}$:

$$\begin{aligned}
(P_{\text{msg}} \times P_{\text{key}})(M_{\Pi} = m) &= P_{\text{msg}}(m), \\
(P_{\text{msg}} \times P_{\text{key}})(C_{\Pi} = c) &= \sum_{m' \in \mathcal{M}, k \in \mathcal{K}} P_{\text{msg}}(m') P_{\text{key}}(k) \llbracket \text{Enc}(m', k) = c \rrbracket, \\
(P_{\text{msg}} \times P_{\text{key}})(M_{\Pi} = m \ \& \ C_{\Pi} = c) &= P_{\text{msg}}(m) \sum_{k \in \mathcal{K}} P_{\text{key}}(k) \llbracket \text{Enc}(m, k) = c \rrbracket,
\end{aligned} \tag{12}$$

where $\llbracket \text{Enc}(m, k) = c \rrbracket := 1$ if $\text{Enc}(m, k) = c$ holds and $\llbracket \text{Enc}(m, k) = c \rrbracket := 0$ otherwise.

Suppose that the random variables M_{Π} and C_{Π} are independent on $U_{\mathcal{M}} \times P_{\text{key}}$. It follows from (12) that

$$\sum_{k \in \mathcal{K}} P_{\text{key}}(k) \llbracket \text{Enc}(m, k) = c \rrbracket = \frac{1}{\#\mathcal{M}} \sum_{m' \in \mathcal{M}, k \in \mathcal{K}} P_{\text{key}}(k) \llbracket \text{Enc}(m', k) = c \rrbracket \tag{13}$$

for every $m \in \mathcal{M}$ and $c \in \mathcal{C}$. Note that the left-hand side of (13) is independent of m . Let P_{msg} be an arbitrary finite probability space in $\mathbb{P}(\mathcal{M})$. For each $m \in \mathcal{M}$ and $c \in \mathcal{C}$ we see that

$$\begin{aligned}
(P_{\text{msg}} \times P_{\text{key}})(C_{\Pi} = c) &= \sum_{m' \in \mathcal{M}} (P_{\text{msg}} \times P_{\text{key}})(M_{\Pi} = m' \ \& \ C_{\Pi} = c) \\
&= \sum_{m' \in \mathcal{M}} P_{\text{msg}}(m') \sum_{k \in \mathcal{K}} P_{\text{key}}(k) \llbracket \text{Enc}(m', k) = c \rrbracket \\
&= \sum_{m' \in \mathcal{M}} P_{\text{msg}}(m') \sum_{k \in \mathcal{K}} P_{\text{key}}(k) \llbracket \text{Enc}(m, k) = c \rrbracket \\
&= \sum_{k \in \mathcal{K}} P_{\text{key}}(k) \llbracket \text{Enc}(m, k) = c \rrbracket,
\end{aligned}$$

where the second and third equalities follow from (12) and (13), respectively. It follows from (12) that the random variables M_Π and C_Π are independent on $P_{\text{msg}} \times P_{\text{key}}$. Since P_{msg} is an arbitrary finite probability space in $\mathbb{P}(\mathcal{M})$, we have that the encryption scheme Π is perfectly secret. This completes the proof. \square

Then, on the one hand, the proof of Theorem 44 is given as follows.

Proof of Theorem 44. Note that $M_\Pi(P_{\text{msg}} \times P_{\text{key}}) = P_{\text{msg}}$ for every $P_{\text{msg}} \in \mathbb{P}(\mathcal{M})$, and $U_{\mathcal{M}}$ is computable. Thus, the theorem follows from Theorems 25 and 39 using Lemma 46. \square

On the other hand, the proof of Theorem 45 is given as follows.

Proof of Theorem 45. Since P_{key} is computable, $C_\Pi(P_{\text{msg}} \times P_{\text{key}})$ is also computable for every computable $P_{\text{msg}} \in \mathbb{P}(\mathcal{M})$. Note also that $U_{\mathcal{M}}$ is computable. Hence, the theorem follows from Theorem 39 using Lemma 46. \square

In the cryptographic community, the notion of perfect secrecy for an encryption scheme is said to imply that even if the eavesdropper has *infinite computing power*, she cannot obtain any information about a message from the corresponding ciphertext. We may interpret the conditions (iv)–(vii) of Theorem 44 as implying this situation in a certain sense since they state that the randomness of a message $M_\Pi(\alpha)$ cannot be reduced even by Martin-Löf tests of unlimited computing power with a complete reference to the corresponding ciphertext $C_\Pi(\alpha)$ as side information.

9.3 Application to quantum mechanics

The notion of probability plays a crucial role in quantum mechanics. It appears in quantum mechanics as the so-called *Born rule*, i.e., the *probability interpretation of the wave function*. In modern mathematics which describes quantum mechanics, however, probability theory means nothing other than measure theory, and therefore any operational characterization of the notion of probability is still missing in quantum mechanics. In this sense, the current form of quantum mechanics is considered to be *imperfect* as a physical theory which must stand on operational means.

As a *major application* of our framework, we can present an *alternative rule to the Born rule* based on the notion of ensemble for the purpose of making quantum mechanics *perfect*. Namely, we can use the notion of ensemble to state the alternative rule for specifying the property of the results of quantum measurements *in an operational way*. We can then present an alternative rule to the Born rule for mixed states based on the notion of ensemble. In particular, we give a precise definition for the notion of mixed state. Finally, we can show that all of the alternative rules for both pure states and mixed states can be derived from a single postulate, called the *principle of typicality*, in a unified manner. We do this from the point of view of the *many-worlds interpretation of quantum mechanics* [10].

The application has been developed in a series of works [21, 22, 23, 24, 25, 26, 27]. A full paper which summarizes the detail of the application is in preparation.

10 Concluding remarks

In this paper we have developed an operational characterization of the notion of probability. As the first step of the research of this line, we have considered only the case of finite probability space, where the sample space is finite, for simplicity. As the next step of the research, it is natural to consider the case of *discrete probability space*, where the sample space is *countably infinite*. Actually, in this case we can develop a framework for the operational characterization of the notion of probability in almost the same manner as the case of finite probability space. The detail is reported in another paper.

Acknowledgments

This work was partially supported by JSPS KAKENHI Grant Numbers 24540142, 15K04981. This work was partially done while the author was visiting the Institute for Mathematical Sciences, National University of Singapore in 2014.

References

- [1] R. B. Ash, *Information Theory*. Dover Publications, Inc., New York, 1990.
- [2] L. Bienvenu, W. Merkle, and A. Nies, Solovay functions and K -triviality, Proceedings of the 28th Symposium on Theoretical Aspects of Computer Science (STACS 2011), pp.452–463, 2011.
- [3] V. Brattka, J. Miller, and A. Nies, “Randomness and differentiability,” preprint, 2012.
- [4] P. Billingsley, *Probability and Measure*, 3rd ed. John Wiley & Sons, Inc., New York, 1995.
- [5] G. J. Chaitin, “A theory of program size formally identical to information theory,” *J. Assoc. Comput. Mach.*, vol. 22, pp. 329–340, 1975.
- [6] G. J. Chaitin, *Algorithmic Information Theory*. Cambridge University Press, Cambridge, 1987.
- [7] A. Church, “On the concept of a random sequence,” *Bulletin of the American Mathematical Society*, vol. 46, pp. 130–135, 1940.
- [8] P. A. M. Dirac, *The Principles of Quantum Mechanics*, 4th ed. Oxford University Press, London, 1958.
- [9] R. G. Downey and D. R. Hirschfeldt, *Algorithmic Randomness and Complexity*. Springer-Verlag, New York, 2010.
- [10] H. Everett, III, ““Relative State” formulation of quantum mechanics,” *Rev. Mod. Phys.*, vol. 29, no. 3, pp. 454–462, 1957.
- [11] O. Goldreich, *Foundations of Cryptography: Volume 1 – Basic Tools*. Cambridge University Press, New York, 2001.

- [12] J. Katz and Y. Lindell, *Introduction to Modern Cryptography*. Chapman & Hall/CRC Press, 2007.
- [13] A. N. Kolmogorov, *Foundations of the theory of probability*, Chelsea Publishing Company, New York, 1950.
- [14] P. Martin-Löf, “The definition of random sequences,” *Information and Control*, vol. 9, pp. 602–619, 1966.
- [15] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information*. Cambridge University Press, Cambridge, 2000.
- [16] A. Nies, *Computability and Randomness*. Oxford University Press, Inc., New York, 2009.
- [17] M. B. Pour-El and J. I. Richards, *Computability in Analysis and Physics*. Perspectives in Mathematical Logic, Springer-Verlag, Berlin, 1989.
- [18] C.-P. Schnorr, “Process complexity and effective random tests,” *J. Comput. System Sci.*, vol. 7, pp. 376–388, 1973.
- [19] C. E. Shannon, “A mathematical theory of communication,” *Bell Syst. Tech. J.*, vol. 27, pt. I, pp. 379–423, 1948; pt. II, pp. 623–656, 1948.
- [20] C. E. Shannon, “Communication theory of secrecy systems,” *Bell Syst. Tech. J.*, vol. 28, pp. 656–715, 1949.
- [21] K. Tadaki, Reformulating quantum mechanics by algorithmic randomness. Presentation at Ninth International Conference on Computability, Complexity and Randomness (CCR 2014), June 9-13, 2014, Institute for Mathematical Sciences, National University of Singapore, Singapore.
- [22] K. Tadaki, A refinement of quantum mechanics by algorithmic randomness. Presentation at Quantum Computation, Quantum Information, and the Exact Sciences (QCOMPINFO2015), January 30-31, 2015, Ludwig-Maximilians-Universität München, Munich, Germany.
- [23] K. Tadaki, “A refinement of quantum mechanics by algorithmic randomness: extended abstract,” *RIMS Kokyuroku* 1952, pp. 112–116, June 2015.
- [24] K. Tadaki, A refinement of quantum mechanics by algorithmic randomness. Proceedings of the Workshop on Informatics 2015 (WiNF 2015), pp. 189–199, December 5, 2015, Meijo University, Nagoya, Japan.
- [25] K. Tadaki, The principle of typicality. Presentation at Eleventh International Conference on Computability, Complexity and Randomness (CCR 2016), January 4-8, 2016, University of Hawaii at Manoa, Honolulu, USA.
- [26] K. Tadaki, A refinement of quantum mechanics by algorithmic randomness. Poster Presentation at 19th Conference on Quantum Information Processing (QIP 2016), January 10-15, 2016, The Banff Centre, Banff, Canada.

- [27] K. Tadaki, A refinement of quantum mechanics by algorithmic randomness. To appear in the Proceedings of the 35th Quantum Information Technology Symposium (QIT35), November 24-25, 2016, High Energy Accelerator Research Organization, Tsukuba, Japan.
- [28] M. van Lambalgen, *Random Sequences*. Ph.D. dissertation, University of Amsterdam, 1987.
- [29] J. Ville, “Étude Critique de la Notion de Collectif,” Monographies des Probabilités. Calcul des Probabilités et ses Applications. Gauthier-Villars, Paris, 1939.
- [30] R. von Mises, *Probability, Statistics and Truth*. Dover Publications, Inc., New York, 1957.
- [31] R. von Mises, *Mathematical Theory of Probability and Statistics*. Academic Press Inc., New York, 1964.
- [32] A. Wald, “Sur la notion de collectif dans la calcul des probabilités,” *Comptes Rendus des Seances de l’Académie des Sciences*, vol. 202, pp. 180–183, 1936.
- [33] A. Wald, “Die Widerspruchsfreiheit des Kollektivbegriffes der Wahrscheinlichkeitsrechnung,” *Ergebnisse eines Mathematischen Kolloquiums*, vol. 8, pp. 38–72, 1937.
- [34] K. Weihrauch, *Computable Analysis*. Springer-Verlag, Berlin, 2000.
- [35] Wikipedia contributors, “Randomness extractor,” Wikipedia, The Free Encyclopedia, http://en.wikipedia.org/wiki/Randomness_extractor (accessed December 17, 2014).